An Examination of the Influence of Task Difficulty on Engagement, Performance and Self-Efficacy Formation within a Computerised Maze Navigation Task

By

Jason Power

B.Tech (Ed)

A thesis submitted to the University of Limerick in fulfilment of the degree of Doctor of Philosophy

Supervisors: Dr. Raymond Lynch and Dr. Oliver McGarr

Submitted to the University of Limerick, January 2017.

List of Figures:

List of Tables:

Abstract

Currently there exists a dearth in the research regarding the link between task difficulty and self-efficacy formation. Within this project, desirable outcomes often linked to self-efficacy, including engagement and performance, are examined through the medium of a digital maze navigation task. Through five rounds of testing the influence of task difficulty, reward and setting were examined relative to engagement and performance. The earlier rounds of the study facilitated an examination of the influence of task variables on engagement rates and performance. Rounds 1 (n=62) and 2 (n=62) examined the influence of task difficulty throughout a school break period and a typical school week respectively, in order to examine whether the influence of task difficulty was consistent across multiple settings. Rounds 3 (n=61) and 4 (n=66) examined the influence and interaction of manipulated difficulty and manipulated reward. Round 5 (n=66) employed the same difficulty settings used in round 1, but also included a 'sources of self-efficacy scale' in order to examine the influence of task difficulty on self-efficacy formation. The results of this study demonstrate a stable relationship between task difficulty, engagement and performance, which have been previously strongly linked to self-efficacy. The final round of this study provides data that supports a link between task difficulty and self-efficacy formation. Altered self-efficacy levels and associated engagement and performance data provide a unique perspective, from which the theoretically reciprocal relationship between self-efficacy and performance can be explored. The results of the study also highlight the need for further research examining the vicarious source and the role of the practice environment in self-efficacy formation.

Declaration

I hereby certify that this material, which I now submit for assessment on the programme leading to the award of Doctor of Philosophy, is entirely my own work and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Signed: _____ Date: _____

Acknowledgements

I would like to dedicate this thesis to my mother, Margaret. Without the countless ways in which you have supported me, big and small, seen and unseen, I could not have made it through this process. You have provided me with opportunities, often at great personal cost, that have opened the doors to a world that I wouldn't have allowed myself to dream of entering only a few years ago. Thank you.

I would like to thank my father, Rodger, for your support and alternative perspective. It is easy to lose sight of what is truly important when life gets hectic. For listening and providing an alternative view, thank you.

I would like to thank my supervisors. To Dr. Raymond Lynch; for your exceptional attention to detail, professionalism and for never flinching when yet another draft was submitted. To Dr. Oliver McGarr; for your support, patience and unique perspectives that drove me to always consider how I would explain my thesis to a stranger in a bar.

I would like to thank my colleagues in DMT and EHS. I would especially like to thank Brian Devitt, Dr. Adrian O Connor, Jeffrey Buckley, Seamus Harrold, Dr. Donal Canty, Dr. Seamus Gordon and Dr. Niall Seery. To Maureen Kendall for making sure I didn't get into too much trouble at staff outings and for her sagely advice regarding gin.

I would like to thank Dr. Patricia Mannix McNamara for her generosity in terms of advice and time.

Finally I would like to thank the cohabitants of the Boulevard of Broken Dreams for the countless cups of coffee, career saving advice, reckless celebrations and unwavering support.

(Liam you may not have my desk or computer).

CHAPTER ONE: INTRODUCTION

Almost four decades have passed since Bandura's seminal paper outlining his theory of self-efficacy (Bandura 1977). Bandura theorised that the beliefs that individuals held relating to their own competence and capabilities had considerable impacts on their ability to self-regulate (Bandura 1986). Self-efficacy refers to an individual's belief in their capability to achieve a positive outcome from a task or activity (Bandura 2012). Self-efficacy as a construct has shown considerable predictive value when examining performance and attainment in diverse settings (Sitzmann and Yeo 2013). There is also a wealth of empirical studies suggesting a positive relationship between self-efficacy and factors relating to satisfaction (Artino 2008, Liaw 2008), achievement (Schunk 2005) and persistence (Martinez 2003). While the positive influence that self-efficacy has on performance has been well established (Pajares 2003), the influence that the relevant task exerts on the formation of self-efficacy has received comparatively little attention (Usher and Pajares 2008). This has resulted in a scarcity of research examining the role of the task. Initially, the current study will examine the impact of task difficulty variation on engagement and performance. Both engagement and performance have been theoretically (Schunk and Pajares 2009) and empirically (Usher and Pajares 2008) linked to self-efficacy yet to date no studies have examined the nature of this relationship. By examining the impact of task difficulty variation on engagement and performance, the current study aims to inform the theoretically proposed reciprocal relationship between performance and self-efficacy. First examining the impact of task difficulty allowed for "sound conceptual specification of the determinants governing performance in a given domain of functioning and the impediments to realizing desired attainments" (Bandura 2012, p.15).

The examination of how task difficulty influences self-efficacy formation has the potential to contribute to self-efficacy theory in a number of ways. The primary contribution lies in the investigation of the impact of task difficulty on the sources self-efficacy. When examined in conjunction with engagement and performance data, the aforementioned theoretical links between self-efficacy, engagement and performance can be scrutinised through empirical means. In addition, the unique design of the study allows for an examination of the theorised reciprocal relationship between performance and self-efficacy (Bandura 1997).

In order to address these points, the selection of task was seen as a crucial consideration in the initial design. A computerised maze navigation task in the form of the classic Pacman arcade game was chosen for a number of reasons. The task required few prerequisite skills or knowledge, it is a highly adaptable piece of software and it presented a uniquely defined domain, which is a fundamental consideration when examining self-efficacy. Informed by the existing literature the following research questions were derived in order to target key areas of interest previously outlined:

1. What is the relationship between task difficulty and engagement in the given task?
2. What is the associated relationship between task difficulty and performance in the given task?
3. What is the relationship between task difficulty, engagement, performance and self-efficacy formation in the given task?

Primarily this study aims to provide empirical data that will allow for a further examination of the influence of task difficulty on engagement, performance and self-

efficacy formation. The design of this study provides a unique perspective from which to examine these relationships, while also facilitating an examination of the posited reciprocal relationship between self-efficacy and performance.

CHAPTER TWO: A REVIEW OF RELEVANT LITERATURE

The following chapter examines self-efficacy in the context of recent research directions, debates within the field and emerging dearths. The evolution of self-efficacy theory is discussed first by comparing existing definitions. From this point, each critical element is discussed from conception to current state. The first section concludes with an examination of the interaction of task difficulty, self-efficacy and related constructs. The latter half of the review focuses on contested aspects of self-efficacy including posited additional sources, domain specificity and measurement of self-efficacy. This culminates in an analysis of the outlined debates in order to identify potential dearths.

Introduction

The "cognitive revolution" (Pervin 1992) of the 1960's saw exponential growth in self-regulation research. This resulted in a wealth of theoretically and empirically informed discussions (Austin and Vancouver 1996). As interest in the area grew exponentially, a proliferation of theories, and components thereof, emerged. Hulleman *et al.* (2010) outlines the negative impact of this proliferation on the further development of self-efficacy theory. In response to concerns of these potential negative impacts, researchers began to examine the common constituent constructs of the more broadly employed self-regulation theories with a view to establishing the level of inter-relatedness between each (Sitzmann and Ely 2011). In this manner commonalities, expressed as constructs, were examined with a view to theoretical alignment, while avoiding distinction based on terminology. "[C]onstructs are the central means we have for connecting the operations used in an experiment to pertinent theory . . . [and] mislabellings often have serious implications for theory" (Shadish *et al.* 2002, p.65).

Sitzmann and Ely (2011) conducted a meta-analysis in order to explore this proliferation. This study based inclusion of the most popular theories on overall citations. The selected theories included Self-efficacy theory (Bandura 1977), Goal Setting theory (Latham and Locke 1979), Control theory (Carver and Scheier 1982), Action Regulation theory (Frese and Zapf 1994), Response Allocation theory (Kanfer and Ackerman 1989), Social Cognition theory (Zimmerman 1990) and Pintrich's 4 phase self-regulation theory (Pintrich *et al.* 2000). Examining each of these theories in detail goes beyond the remit of this review. However, it is worth examining the commonalities and origin dates of each (see Table 1.). The identified constructs were classified into three groups including Regulatory agent, Regulatory mechanisms and Regulatory appraisals. For a complete description of this division process and inter-rater procedure see Sitzmann and Ely (2011).

Table 1. Constructs Contained within Self-Regulation Theories (Sitzmann and Ely 2011, p.424)

Construct	Bandura (1977, 1997, 1991)	Carver & Scheier (1981, 1990, 2000)	Frese & Zapf (1994); Hacker (1985)	Kanfer & Ackerman (1989)	Locke & Latham (1984, 1990, 2002)	Pintrich (2000)	Zimmerman (1990, 1996, 1998, 2000)
Regulatory agent							
Goal level	X	X	X	X	X	X	X
Regulatory mechanisms							
Planning		X	X		X	X	X
Monitoring		X	X	X		X	X
Metacognition			X	X		X	X
Attention		X		X	X		X
Learning strategies						X	X
Persistence	X	X	X	X	X	X	X
Time management						X	X
Environmental structuring						X	X
Help seeking			X			X	X
Motivation	X	X	X	X	X	X	X
Emotion control				X		X	
Effort	X	X	X	X	X	X	X
Regulatory appraisals							
Self-evaluation		X		X		X	X
Attributions	X	X				X	X
Self-efficacy	X	X		X	X	X	X

Note. An *X* denotes the theory suggests that the construct is a component of self-regulation.

Table 1 outlines the inclusion of constructs in the most popular self-regulation theories. A considerable inter-rater agreement level was achieved (.89) (Sitzmann and Ely 2011, p.423) and the study suggested that the strength of the construct affect on desired outcomes was significantly related to the amount of theories in which it

was present (r = .48) (Sitzmann and Ely 2011). This would suggest that self-efficacy, which is included in six out of the seven theories of self-regulation previously outlined, has a greater predictive value when compared with constructs that appear in fewer theories (see Table 1.) It is worth noting that five of the six constructs attributed to Bandura's (1997) theory are also the five most potent constructs.

> "Meta-analytic findings revealed that the majority of self-regulatory processes have moderate to strong relationships with one another, suggesting that the processes are highly interrelated. Additionally, examining the intercorrelations between self-regulation constructs suggests that there is measurement overlap in the assessment of some constructs"
>
> (Sitzmann and Ely 2011, p.438)

Acknowledging the expansive nature of self-regulation theories (Burnette *et al.* 2013), this review will focus on the impact of the constructs of self-efficacy that could be most susceptible to influence from manipulated task properties. The rationale for selecting self-efficacy, as opposed to any of the other popular self-regulation theories listed previously, is based on two primary arguments. Firstly, self-efficacy significantly correlates with the most popular current self-regulation theories, while being comparatively constrained in its theoretical scope (Schunk and Pajares 2009). Secondly, self-efficacy has demonstrated greater explanatory and predictive value when the task is well defined (Stajkovic and Luthans 1998, Haddad and Taleb 2016, Honicke and Broadbent 2016). Having established a rationale for the selection of self-efficacy it is worth examining the conception of the theory itself.

The Origins of Self-efficacy Theory

Social Cognitive Theory posits that individuals "function as contributors to their own motivation, behaviour, and development within a network of reciprocally interacting influences" (Bandura 1999, p. 169), as opposed to organisms reacting to stimuli in

their environment. In essence, it explores the influence of behavioural, cognitive and environmental factors on the development of behaviour and motivation. It is theoretically argued that these factors are inseparable and influence each other in a reciprocal manner; this is referred to as Triadic Reciprocal Determinism by Bandura (1989). However, few studies have provided empirical data that supports this model (Williams and Williams 2010). This triadic model is positioned as a component of the broader Social Cognitive Theory. It is important to note that it is not claimed that each factor is equal; nor that they exert influence concurrently (Wood and Bandura 1989). In addition, the resultant determinism is described as probabilistic and acknowledges the potential impact of unidentified elements on any of the behavioural, cognitive or environmental factors (Bandura 1983).

Bandura's social cognitive theory (Bandura 1986) proposed that individuals perform or behave in a manner that is influenced by internal mechanisms such as self-reflection, cognition and vicariousness. When considered in terms of self-efficacy these factors are said to be task orientated rather than general environmental factors. Bandura (1997) suggests that what an individual believes, rather than what is objectively true, is a stronger indicator of performance, motivation and well-being (Tschannen-Moran and Hoy 2007). If an individual does not believe that their actions can have a meaningful and positive result they have no incentive to attempt said action. It is for these reasons that an individual's belief can often be a better indicator of future performance than actual ability in a given field (Sitzmann and Yeo 2013). Self-efficacy can affect the manner in which an individual negotiates problems, both in cognitive constructions of solutions and analyses of requirements, and in emotional response (Bandura 1997, Martinez 2003).

Bandura (1986, 1997) theorises that self-efficacy develops as an individual interprets information from four sources; *Mastery Experience*, *Vicarious Experience*, *Social Persuasion,* and *Physiological States*. The impact of each source varies by domain and is also governed by various cognitive processing strategies that the individual may employ (Joo *et al.* 2013). While mastery is generally acknowledged to be the predominant source, the contributions of the other sources differ by domain and is the topic of ongoing investigations (Anderson and Betz 2001, Caraway *et al.* 2003, Britner and Pajares 2006, Bandura 2012). A review by Usher and Pajares (2008) outlines this complex and dynamic relationship between source dominance. In line with Bandura's (1986, 1997) original hypotheses; mastery proves to be the most consistent and reliable predictor of self-efficacy (Hampton 1998, Wang and Newlin 2002, Britner and Pajares 2006, Caprara *et al.* 2008). Each source is examined in detail in a later section. Prior to this, the development and definition of self-efficacy are examined.

Defining Self-efficacy

Self-efficacy was originally proposed as a component of Bandura's Social Learning Theory, which later evolved into what is now referred to as Social Cognitive Theory (Ashford and LeCroy 2000). Bandura's (1986) work outlining his Social Cognitive Theory expanded upon the influences of self-efficacy to include the effort individuals expend, choices made, levels of anxiety, levels of persistence and perseverance as individuals engage in the wide variety of tasks that make up their daily lives. Self-efficacy was largely a response to what Bandura perceived as the inadequacies of behaviourism and psychoanalysis, which failed to acknowledge the role of cognition in the formation of motivation (Bandura 1977). Bandura (1997) suggests that what an individual believes, rather than what is objectively true, is a

stronger indicator of performance, motivation and well-being. If an individual does not believe that their actions can have a meaningful and positive result they have no incentive to attempt an action. This has further implications for perseverance in the face of adversary (Hamilton 2011). It is for these reasons that an individual's belief can often be a better indicator of future performance than actual ability in a given field (Bandura 1997). Self-efficacy can also affect the manner in which an individual negotiates problems, both in cognitive constructions of solutions and analysis of requirements, and emotional response (Bandura 1997, Honicke and Broadbent 2016). Low levels of self-efficacy can lead to anxiety as well as numerous other negative and potentially performance debilitating feelings, conversely high levels of self-efficacy can enhance confidence and task performance (Magaletta and Oliver 1999, Lee 2015).

While self-efficacy should be considered as a factor in task performance; it is worth considering the construct relative to other self-concepts. Recent findings suggest that self-concepts that are domain specific present significant similarities and correlations to self-efficacy (Parker *et al.* 2014). However, self-efficacy has demonstrated greater predictive value where the task outcome is relatively predictable in nature and the overall task requirements are familiar to the individual (Pajares and Miller 1995). This highlights the influence of the nature of the task in self-efficacy formation. However, when examining self-efficacy formation it is necessary to first examine the ontogenetic development of self-efficacy.

Ontogenetic Development of Self-efficacy

The development of self-efficacy begins in early infancy and is influenced by parents or other key figures in the child's developmental period (Bandura 1997). The environment that the child develops in provides key stimuli. An environment that

provides accomplishable challenges and recognition of success enhances the development of cognition and self-efficacy (Bandura 1997). This is also fundamental to Flow Theory and multiple other competence based theories (White 1959, Bandura 1982, Deci and Ryan 1985, Harter 1996). While initially the child will not directly influence its environment, over time exhibitions of curiosity or engagement with a stimulating environment can in turn influence the self-efficacy of the carer in their ability to aid the development of the child (Bandura 1997). This leads to a mutually influential relationship, which demonstrates the reciprocal nature of the theory. Through increased self-efficacy of the care giver the environment continues to evolve to meet the child's needs. As the child engages with new tasks or challenges they can acquire knowledge and skills at an increased rate (Bandura 1997). Miller and Meece (2001) indicate that by providing an environment that allows for exploration, encourages curiosity and facilitates engagement in play related activities it is possible to accelerate self-efficacy development (Pajares and Urdan 2006). Bandura (1997) proposed that parents who provided the opportunity and freedom to engage with challenging tasks tended to have more self-efficacious children than those that did not. The role of the parent in selecting the difficulty of this task is critical for self-efficacy development. Bandura (1997) argues that a task that is too difficult can result in a reluctance to engage and reduced persistence. Negative mastery experiences, arising from successive failures within a domain, can lead to a sense of inefficacy resulting in disengagement and problematic self-handicapping behaviour (Bandura *et al.* 1999). This study aims to examine the influence of difficulty on engagement and mastery experience, but will also examine the potential influence of difficulty on the other three sources: social persuasions, physiological state and vicarious experience.

The vicarious source plays a crucial role in early self-efficacy development. Through modelling persistence and engagement with challenging activities, parents allow their children to develop self-efficacy vicariously. Modelling provides an opportunity to observe proficiencies that the individual wishes to achieve (Bandura 1995). Vicarious development can also be encouraged through peer interaction, but only if the individual believes that his or her peers are of an equitable ability level. If the individual believes that their peers have significantly higher abilities in a certain field then this effect is negated (Bandura 1982). Observing a peer engage with a challenging task can lead to raised self-efficacy levels for the same, or similar, task (Schunk 1987). Conversely observing peers that fail to complete a task can have a predictably negative influence. Peer modelling of this nature is particularly strong in adolescents (Holden *et al.* 1990), which raises many questions relating to the environmental and peer impacts on self-efficacy. The peer group that a child belongs to can have significant effects on self-efficacy and motivation (Britner and Pajares 2006). A normalising effect can occur over an extended period of time where the individual's beliefs regarding their own abilities tend to sway towards the group belief (Kinderman *et al.* 1996). Steinberg et al. (1996) indicated that peer pressure peaks between the ages of 12 to 16. During this time the influence that a parent has in their child's self-efficacy development lessens and is replaced by peer influences (Steinberg *et al.* 1996). This time is also subject to a number of influences which are viewed as being predominantly damaging to self-efficacy and competence based beliefs (Caraway *et al.* 2003).

Pintrich and Schunk (2002) suggest that self-efficacy and motivation decline throughout an individual's educational experience. This decline has been linked to an increase in grading, less individual involvement with teachers, ability grouping, peer

competition and various stresses involved in major examinations. Interestingly the perceived efficacy levels of teachers has an indirect link to student self-efficacy (Pintrich and Schunk 2002). Woolfolk (1990) suggests that teachers with a low level of self-efficacy tend to view pupils as being less motivated and interested, thus they favour a more rigid form of classroom management that can reduce opportunities for pupils to engage in challenging activities. This, in turn, lowers pupils' efficacy levels and the negative cycle continues. This manner of instruction typically relies on extrinsic style motivators and sanctions.

The studies mentioned within this section focus on the optimum environments and influences that lead to increased self-efficacy. This is based on the link between increased self-efficacy and increased performance in a range of diverse fields or domains (Sitzmann and Yeo 2013). This simplistic unidirectional model of causation is currently the subject of considerable debate within the field (Williams and Williams 2010). Sitzmann and Yeo (2013) suggest that the relationship between self-efficacy and performance is more nuanced and that many research designs are not optimised to examine this relationship. In order to robustly examine the nature of the relationship between these factors, it is first necessary to examine the four sources of self-efficacy as theorised by Bandura (1986).

Sources of Self-efficacy

Bandura (1986, 1997) theorises that self-efficacy develops as an individual interprets information from four sources; *mastery experience*, *vicarious experience*, *verbal and social persuasions* and *emotional and physiological states*. The most influential of these four sources is *mastery experience* (Bandura 1997). This is the result of an individual engaging in a task and achieving what they perceive to be a positive

outcome[1]. This experience of mastery leads to increased interpretations of their own capabilities in similar tasks (van Daal *et al.* 2014, Zelenak 2015). Successive mastery experiences have a predictably positive affect on self-efficacy. This is not to say that isolated failure will halt the development of this source. Rather an individual will typically form their own perception of mastery over multiple engagements and a period of time. This allows for isolated failure (Caprara *et al.* 2008). Indeed it has been suggested that mastery experiences that involved overcoming significant obstacles result in greater, and more resilient, levels of self-efficacy than those that are easily achieved (Holden *et al.* 1990, Bandura 1997). However, mastery is not the sole contributor to a fully developed sense of self-efficacy. Mastery must be interpreted in conjunction with a myriad of other factors including perceived task difficulty, help received, effort spent as well as previously held self-beliefs (Usher and Pajares 2008). Effort can often be used to assess one's own capability in the absence of a more objective measure. This could potentially undermine a mastery experience if the individual perceives that the amount of effort required to achieve a successful outcome in a task is far greater than for a comparable peer resulting in a deterioration of efficacy beliefs (Joo *et al.* 2013). Conversely, a positive outcome where the individual perceives that they exerted less effort than a comparable peer, with an equitable or more desirable result, can lead to enhanced efficacy beliefs. Similarly, a positive result achieved with the aid of others may negate positive efficacy developments as the success could potentially be attributed to another resulting in a loss of mastery experience (Schunk 1995).

[1] Within this review 'mastery experience' is defined as an individual's positive perception of performance. This is not directly equatable to 'performance' which is typically expressed as a metric, such as a percentage or grade. This aligns with the tenets of self-efficacy theory; the negative implications of blurring this distinction are outlined later in this chapter (Bandura, 2012).

Vicarious experience is an additional source of information which aids in the development of self-efficacy. This occurs when an individual observes another completing a similar task. For example observing a classmate complete a task that the individual has previously not encountered could enhance their sense of efficacy for the given task (Schunk 1987, Bandura 1997). The similarity of the peer relative to the individual also has an influence. The more equatable the individual believes the observed party to be; the greater the effect. Bandura (1997) suggests that this source plays a lesser role in self-efficacy formation when compared to mastery as a source. This is supported by recent studies in the area (Parker *et al.* 2014, Lee 2015, Geitz *et al.* 2016). However, when individuals have not previously encountered a particular activity, and as a result are unsure of their own ability, vicarious experience can assume source dominance due to lower mastery source influence (Usher and Pajares 2008).

Social persuasions play a key role in the development of self-efficacy. Feedback, both verbal and non-verbal, can encourage resilience and nurture confidence (van Dinther *et al.* 2011). Encouragement of this nature enhances an individual's confidence while ensuring the predicted success is achievable (Williams and Williams 2010). Negative persuasions operate in a predictably opposite manner with some authors maintaining that negative persuasions are far more likely to leave a lasting effect when compared to the impact duration of positive persuasions (Usher and Pajares 2008). Feedback that aids the individual to identify areas to improve, and highlights past improvements rather than outlining how far the individual must still progress, has a far greater impact on self-efficacy (Bandura 1997).

Individuals also draw information regarding their capabilities from their own *Physiological State*. Sensations such as anxiety, fatigue or stress will have a

predictably negative influence on this interpretation of source information (Bandura 1997). Over time individuals learn to interpret these physiological and emotional cues from interactions with diverse tasks and situations. For example, an individual who experiences negative physiological states such as anxiety when faced with public speaking may well perform poorly in the given task leading to lower levels of self-efficacy. This cyclical nature may lead to increased negative physiological cues in future public speaking engagements causing a further deterioration of the individual's perceived self-efficacy relating to public speaking (Usher and Pajares 2008). This highlights various aspects of the task, such as difficulty level, which could potentially induce a negative physiological state thus lowering self-efficacy (Orvis *et al.* 2008). It should be noted that the interpretation of physiological information determines the negative or positive orientation of the source. An individual may interpret an elevated pulse rate as negative, leading to a sense of dread. Conversely, another may interpret the same physiological cue as positive. This may lead to a sense of excitement or arousal (Bandura 2012). As with all the previous sources listed, the individual's perception is the critical factor that defines the nature of the influence of each source. When examining the role of each source in the formation of self-efficacy it is necessary to also consider the relationships between these sources.

Dynamic Source Influence

By analysing information from the four sources outlined previously individuals create their own sense of self-efficacy relative to a given task or discipline. The impact of each source varies by domain and is also governed by various cognitive processing strategies that the individual may employ (Schunk and Pajares 2009). While mastery is generally acknowledged to be the predominant source, the

contributions of the other sources differ by domain and is the topic of ongoing investigations (Anderson and Betz 2001, Caraway *et al.* 2003, Britner and Pajares 2006, Bandura 2012). A review by Usher and Pajares (2008) outlines this complex and dynamic relationship between source dominance. However, in line with Bandura's (1986, 1997) original hypotheses, mastery continues to be the most consistent and reliable predictor of self-efficacy (Hampton 1998, Wang and Newlin 2002, Britner and Pajares 2006, Caprara *et al.* 2008).

While mastery has shown the greatest predictive value it has also been the most frequently misrepresented source (Usher and Pajares 2008). Matsui *et al.* (1990) investigated mathematics self-efficacy in a high school environment in Japan. This study used students past math grades to represent mastery experience. This is an example of misinterpretation of the original theory and the associated concerns of Bandura (2012) discussed previously. Prior math grades are unsuitable as estimates of mastery as it does not account for students' interpretations of grades received (Bandura 2006, Pajares and Urdan 2006). A common analogy is again based within the realm of mathematics. If we imagine how two students receive a B grade, for one student this may represent a positive outcome. For the second student, who is accustomed to A grades, this is perceived as a negative outcome. Bandura suggests that "the same level of performance success may raise, leave unaffected, or lower perceived self-efficacy depending on how various personal and situational contributions are interpreted and weighted" (Bandura 1997, p. 81).

The other sources have been reported to generally have lesser predictive value (Lent *et al.* 1991, Lopez and Lent 1992, Hampton 1998). In addition inconsistent factor models, even within similar domains, are cause for concern (Matsui *et al.* 1990, Lent *et al.* 1991, Anderson and Betz 2001, Usher and Pajares 2006). Britner and Pajares

(2006) posit that these inconsistencies can be attributed to differences in methodological decisions. Many of the studies discussed here utilised hierarchical regression or stepwise models which require variables to be entered in order of influence or relative potency (Matsui *et al.* 1990, Lent *et al.* 1991, Lopez and Lent 1992, Hampton 1998). This is problematic due to a lack of original theoretical support. Although Bandura outlined mastery as the predominant source he did not specify the ranking or relative influence of the other sources (Bandura 1986, Bandura 1997). While the lack of a prescribed hierarchy is not problematic in itself, the nature of stepwise, or hierarchical regression, creates difficulties when trying to ascertain the influence of each source across studies. A review by Usher and Pajares (2008) found that mastery has consistently resulted in higher correlations to self-efficacy with *r* values ranging from .29 to .67 (median = .58). This was greater than any other source in the listed studies and was notably the only consistently significant source (Usher and Pajares 2008). While few studies challenge this predictive relationship, it is worth noting that Gainor and Lent (1998) found no significant correlation with self-efficacy. However, they did note a correlation with social persuasions ($r = 0.70$). This study examined the self-efficacy beliefs of black undergraduates and found that social persuasions did predict self-efficacy. This led the authors to posit that mastery and social persuasions, in this instance, likely drew from the same source of variance. The identification of this common variance source went beyond the remit of the study, but was highlighted as a potential target of further study (Gainor and Lent 1998). Collectively these studies highlight the dynamic nature of source influence on self-efficacy formation. It is essential that researchers are cognisant of this dynamic structure when designing studies in order to provide data that facilitates a more thorough examination of these influences.

Relationship between Self-Efficacy and Performance

The relationship between performance and self-efficacy is governed by the knowledge of the required task; as such the nature of the task bears a considerable influence on the nature of the self-efficacy in question. Schunk (1989) highlights that the relationship between self-efficacy and performance is maximised when the participant has sufficient knowledge of the skills required to successfully complete the task, this is referred to as 'self-efficacy for performance' (Schunk 1989). Where tasks are ill-defined and the participant has less knowledge of the overall requirements their reported self-efficacy is referred to as 'self-efficacy for learning' (Schunk 1989).

The majority of quantitative studies examining self-efficacy and performance have been based in the domain of mathematics (Lent *et al.* 1991, Pajares and Miller 1995, Lent *et al.* 1996, Gérard 1998, Stevens *et al.* 2006, Boston and Smith 2009, Usher and Pajares 2009, Hamilton 2011, Parker *et al.* 2014). Williams and Williams (2010) note that although self-efficacy is examined in terms of its influence on performance, none of the studies listed above explicitly model the reciprocal relationship between the two factors as hypothesised by Bandura (1978). Williams and Williams (2010) suggest that the reason for a lack of empirical evidence examining this reciprocal relationship was due, in part, to the fact that "recursive statistical models […] forced the investigators to assume one position or the other in order to be able to estimate the model" (Williams and Williams 2010, p.456). The restrictive models used typically forced the user to assume one of two possible conceptual positions; a self-enhancement position which suggests that self-beliefs enhance performance, or a skill development position which views self-beliefs as merely a product of performance (Sitzmann and Yeo 2013, p.560). Ignoring the reciprocal underpinnings

of the model has been judged by some to be methodologically and theoretically problematic (Byrne 1996, Marsh and Craven 2006). In addition, it is indicative of a dearth in the research that allows for the examination of the influence of performance on self-efficacy formation in a longitudinal fashion. As noted by Bandura "performance situations contain constellations of factors that convey efficacy information" (Bandura 1997, .p85). The current study aims to address this dearth by employing a novel design that allows for analysis of performance and engagement relative to self-efficacy development over a practice period.

Stajkovic and Luthans (1998) meta-analysis (114 studies, k = 157, N = 21,616) results indicate that self-efficacy increased performance by 28%. This is an example of the previous tendency to focus on a unidirectional enhancement that fails to examine either a performance enhancing effect on self-efficacy or the nature of the relationship between performance and self-efficacy. The authors note that this is a greater effect than those reported in similar meta-analysis studies examining the impact of goal setting, feedback interventions or behavioural interventions (Stajkovic and Luthans 1998). A more recent meta-analysis by Sitzmann and Ely (2011) (k = 430, N = 90,380) suggests a 93% positive reported correlations between self-efficacy and performance. However, these results are typically reported as self-efficacy's impact on performance. Beattie *et al.* (2011) question the nature of this relationship. Their findings suggest that rather than self-efficacy primarily acting as a driver of future performance, it is primarily a product of past performance (Beattie *et al.* 2011). It should be noted when examining this relationship Beattie *et al.* (2011) compare performance metrics (objective performance scores) to composite self-efficacy scores. This provides limited opportunities to examine the manner in which performance and self-efficacy interacted.

A further meta-analytical study (Sitzmann and Yeo 2013) examining the nature of this relationship reached a similar conclusion to that of Beattie *et al.* (2011). Sitzman and Yeo's (2013) study suggests that the impact self-efficacy had on performance ranged from null to moderate (ρ ranged from -.02 to .33). A stronger influence was observed when considering the effect of past performance on self-efficacy formation (ρ ranged from .18 to .52). This leads the authors to conclude that the effect that past performance has on self-efficacy is greater than the comparable effect of self-efficacy on performance (Sitzmann and Yeo 2013). The author's also note the lack of studies examining the impact of performance on self-efficacy. This argument resonates with elements of self-efficacy theory. Mastery experience as the primary source of self-efficacy (Parker *et al.* 2014, Lee 2015) is logically influenced by the nature of the task. Yet none of the studies listed within Sitzmann and Yeo (2013) explicitly examine the influence of task variables on self-efficacy formation. If mastery experience is interpreted as previous positive achievement within the same domain or task (Bandura 1977), then the difficulty level of that task should play a role in mastery experience, and as a result self-efficacy formation.

Defining Difficulty

When considering the impact of task difficulty it is first necessary to examine the various definitions of difficulty. Fodor (1974) highlights the problems that arise from multiple definitions of "difficulty" and provides a concise overview of the need for either an agreed definition or an explicit selection of a definition by the author in studies that employ difficulty as a variable. Gilbert *et al.* (2012) echoes this concern and provides an overview of the more common definitions of difficulty.

The first of these definitions hinges on whether the task is accomplishable. Tied to this is the developmental perspective, which examines whether the rate of task

completion relative to the participants developmental stage. Using this definition the majority of younger participants in a lesser developmental stage would not be able to complete the task, while the majority of a group at greater developmental stage should be able to complete the task. However, this has limited discriminant validity when examining healthy homogenous populations. In order to overcome this limitation a series of tasks are often presented sequentially with increasing difficulty levels and scored by percentage completion (Gilbert *et al.* 2012).

A second definition of task difficulty is based on the concept of dual task interference. Using this definition the difficulty of the task in question is judged by the degree of negative influence it exerts on a parallel task that a participant must complete while completing the primary task. This definition presents multiple problems relating to the direction of influence between the primary task and the parallel task (Gilbert *et al.* 2012). In addition this definition has proved problematic due to unidentifiable specific resource requirements relating to attention (Allport *et al.* 1972).

Within the field of neuropsychology, task difficulty is often defined based on response time, dual task interference or a combination of both (Gilbert *et al.* 2012). These tasks are typically simplistic in design and, due to the restrictions of fMRI technology, are rarely longitudinal in nature (Gilbert *et al.* 2012). The current study subscribes to this definition of task difficulty and this is reflected in the maze navigation task employed by the study. In creating varying difficulty versions of the maze navigation task the speed of negative characters was manipulated in a positive and negative manner. These manipulations require decreased response time in order to progress within the task. A more complete description of the maze navigation

task, manipulations and its development, is presented in the subsequent methodology chapter.

Task Difficulty and Performance

Kumar and Jagacinski (2011) examined the interaction between performance based goals and task difficulty in the context of goal type appropriation. The authors examined the behaviours and goals of participants using Nicholls (1984) framework. This framework encompassed the behaviours and goals of participants when encountering difficulty. According to Nicholls (1984), as individuals encounter increasing levels of difficulty in a task their perceived ability should decline. As an individual's perception of ability declines it is likely that they will shift from performance based goals to performance avoidance goals. This change is to protect the self-image of the individual. Further difficulty increases at this stage would lead to work-avoidance goals where the individual would aim to stop all participation in the task (Maehr and Zusho 2009). Individuals with increased self-efficacy, relative to the task in question, would be slower in their transfer from performance to performance avoidance, and ultimately work avoidance goals (Jiang *et al.* 2014, Geitz *et al.* 2016, Haddad and Taleb 2016).

This change in emotional response relative to a task can be referred to as task valence (Silvestrini and Gendolla 2009). Attractiveness (positive valence) to a task aligns with performance goal orientations (Kumar and Jagacinski 2011) and is also related to a heightened sense of self-efficacy (Chiou and Wan 2007). Individuals that demonstrate positive task valence typically demonstrate increased levels of task engagement, lower attrition and greater overall performance (Feather 1992, Silvestrini and Gendolla 2009). The positive behaviours observed in individuals who

demonstrated positive task valence align with behaviour associated with increased self-efficacy (Chiou and Wan 2007).

Conversely, those that demonstrate negative task valence typically demonstrate non-desirable behaviour similar to those who have adopted performance avoidance or work avoidance goals (Silvestrini and Gendolla 2009). Negative valence is also a corollary of reduced engagement, retention and performance (Silvestrini and Gendolla 2009). From a theoretical and empirical perspective this aligns with individuals who demonstrate lower levels of self-efficacy (Joo *et al.* 2013, Zelenak 2015). The common resultant of negative task valence, performance avoidance goals and low self-efficacy is a negative outcome for a task (Liem *et al.* 2008, Silvestrini and Gendolla 2009, van Daal *et al.* 2014). This highlights the importance of task design, and the impact of task properties, and is especially true of task difficulty. These closely related theories suggest that excessive task difficulty would have detrimental impacts on perceptions of ability and would result in disengagement from the task at an early stage (Gilbert *et al.* 2012). This has obvious negative connotations for attrition and overall performance. It is however necessary to acknowledge that factors that influence performance should not be considered in isolation. The environment in which the individual exists impacts a variety of factors that affect self-regulation, not least of which is the social influence (Schunk and Pajares 2009).

Task Difficulty and Social Influence

The impact of social norms on performance in various domains has been a popular area of investigation within the field of social psychology dating back almost as far as the field itself. Sherif's (1936) seminal work examining social influence pays considerable attention to the impact of task difficulty. In this study participants were

asked to estimate movement of a point of light. The point of light was in reality stable. As participants were exposed to incorrect estimations, they altered their own estimations of the inherent difficulty of the task. From this foundational work multiple well renowned studies reached a similar conclusion; individuals become more receptive to social influence as the perceptive difficulty of a task increases (Klein 1972, Nordholm 1975, Baron *et al.* 1996, Rosander and Eriksson 2012). One of the more noteworthy studies to emerge from this line of research is that of Janis (1972) who examined the impact of difficulty on group behaviour. The paper suggests that a task that is perceived to be more difficult not only results in greater social influence, but that the resultant decision making of a group typically displays ill-conceived and objectively poor decision making abilities. In addition the perceived increased difficulty also increased susceptibility to stress.

Brown and Cialdini (2015) explore the question of why higher task difficulty increases the influence of social influence. The paper explores the influence of a desire to act in a socially defined "correct" manner in terms of evolutionary theory. Ultimately the paper argues that as the perceived difficulty of a task increases, the individual's perception that they may act in an "incorrect" manner also increases. This causes the individual to weigh social influence to a greater degree. Brown and Cialdini (2015) deduce that this prosocial behaviour leads to lowered perceptions of stress and negative emotions. Although the papers previously discussed in this section provide a strong theoretical support for the influence of task difficulty on an individual's susceptibility to social influence, a considerable amount of research provides either contrary evidence or findings that suggest a null effect. Seaborne (1962) suggests that the relationship between task difficulty and social influence is not as direct, nor simplistic, as some had previously suggested. Seaborne noted that

the relationship between social influence and conformity was confounded by the ambiguity of the task. Stasser and Dietz-Uhler (2008) conducted a comprehensive review of this area and reached a similar conclusion. Both sources highlight the impact of the task itself. The very nature of the task is deemed to have a considerable impact on social influence, with tasks demonstrating ambiguous definitions leading to confounding results (Stasser and Dietz-Uhler 2008). The nature of the task, and associated performance indicators, has been a source of similar concerns within self-efficacy theory (Geitz *et al.* 2016, Haddad and Taleb 2016). Brown and Cialdini (2015) posit that one potential cause for this lack of consistency in experimental settings is due to the manner with which an individual engages with social persuasion. As previously discussed in terms of prosocial behaviour (Brown and Cialdini 2015), some individuals may wish to be perceived as making the socially defined correct decision. However, some may wish to project independence, or protect another self-concept, that runs counter to this typical interaction (Cialdini and Trost 1998). When the wider research base is considered it suggests that task difficulty plays a role in the manner in which an individual is affected by social influence. This relationship can prove difficult to isolate in experimental conditions due to influence from various motives and self-concepts (Wood 2000). Baron *et al.* (1996) suggests that one such motive could be linked to perceived task value or importance.

Subjective Task Value

Task value is a critical factor which governs the engagement levels and retention of an individual when completing a task. *Subjective Task Value* (STV) (Wigfield and Eccles 1992) was originally developed as an expectancy-model theory specifically

addressing a socio-cultural phenomenon in achievement related choices[2]. The development of this theory was heavily influenced by numerous previous theories and is in essence a hybrid of more focused motivation, achievement and value based theories (Battle 1965, Feather 1992, Eccles 2005). As such, it serves as a useful structuring device for the review of topics relevant to these areas. STV theory assumes that the task presents an inherent value to an individual. The value is a quality that is attached to this task and contributes to the probability of the individual selecting or engaging with that task (Eccles-Parsons *et al.* 1983). Eccles and Wigfield (1995) suggest that an individual could have relatively high self-efficacy beliefs relating to a task, but chose not to engage in said task because it has little subjective value. Eccles and Wigfield (1995) suggest that task value is a product of subjective beliefs related to the task, the perceived purpose of the task and self-defined goals.

Eccles and Wigfield (1995) posit that task value consists of three types of value; interest, importance and utility. Interest relates to the inherent enjoyment that an individual experiences when completing the task. The importance value is based on the significance that an individual places upon successful completion of a task and as such acts as an affirmation of their own skill level, increasing self-image, if the outcome is positive (Eccles 2005). Equally a negative outcome will result in damage to the self-image. This value is closely related to the work of Battle (1965) who further defined attainment value in his own theory. Battle (1965) suggests that the importance of a task increases or decreases depending on how the individual perceives that task in relation to their core social and personal identities. This theory

[2] Subjective Task Value is a consolidated theory of considerable size, for a more complete summary of the theory see Eccles, J. S. (2005) 'Subjective task value and the Eccles et al. model of achievement-related choices' in Elliot, A. J. and Dweck, C., eds., *Handbook of competence and motivation*, New York: The Guilford Press, 105-121.

in turn is similar in many elements to that of Skinner *et al.* (1990), which focuses more on social aspects of attainment theory. The social aspects are more fully explored and defined in Connell and Wellborn (1991), which adds two basic needs as attainment value prerequisites: 1) the need to feel that what one does is fundamentally important to one's social group. 2) the need to feel respected and valued by one's social group. The second is particularly relevant to variable difficulty tasks given the risk of displaying inferiority to one's social group. This in turn presents a potential degrading effect on one's self-image and related elements previously discussed. This need to display a level of competence, and the social risks it entails, has been widely examined in many of the theories previously discussed. This affirmation of ability bears considerable parallels to mastery experience cited as the predominant source of self-efficacy (Zelenak 2015).

The final type of constituent value is that of utility. Utility is described as the perceived worth of a task relative to an individual's future goals (Wolters *et al.* 1996). Utility value is dictated by a perceived alignment with an individual's long or short term goals. If the successful completion of a task is perceived to be of benefit to the performance of everyday activities, or for the individual's long term ambitions, then the task is said to be of high utility value (Eccles *et al.* 2005). There exists a considerable body of research that suggests that utility value is positively associated with a range of desirable outcomes such as effort (Cole *et al.* 2008), interest (Hulleman *et al.* 2008) and performance (Hulleman and Harackiewicz 2009). While a relatively arbitrary task, such as the maze navigation task used in this study, could be said to have comparatively low utility value, it is important to consider the task outcome relative to the individual's desires. Bong (2001) notes that utility value for younger students, such as the participants in this study, can often be directed

towards pleasing authority figures or allowing for favourable comparisons with peers.

Eccles and Wigfield (1995) posit that the three types of task values act in a collective fashion and when considered as a whole can be labelled as positive task valence. Multiple studies demonstrate the links between elevated task value and increased engagement, retention and self-efficacy (Wang and Newlin 2002, Bunn 2004, Wang *et al.* 2013, Lee 2015). Chiu and Wang (2008) examined the elements of subjective task value, as well as other previously outlined constructs with regard to continuance intention. This is framed as a key requirement of any self-directed task.

The results of this study are particularly relevant given the similarities between elements of the model and the sources of self-efficacy previously examined. It is worth noting that intrinsic value (playfulness) displays the greatest explanatory value for continuance intention (Chiu and Wang 2008). Intrinsic value is defined by the authors as "the extent to which an activity is perceived to be personally enjoyable. According to self-determination theory, learners are self-determining and intrinsically motivated […] when they are interested in or enjoying doing it." (Chiu and Wang 2008, p.196). This is particularly relevant to the design of the current study as the selection of the task was seen as a critical component. If the task was not seen to be potentially enjoyable, then continuance intention would presumably have been low as a result. Given a key premise of the study is to examine the influence of task properties on engagement, performance and self-efficacy; a task that was not conducive to continuance intention would have reduced the value of the data when examining task difficulty's influence. Ultimately continuance intention outlined above can be considered a self-regulatory mechanism. While its relevance to this

study and its links to self-efficacy theory are clear, it is worth examining additional posited links to goal theory (Chiu and Wang 2008).

Self-Efficacy, Goals and Goal Orientation

Komarraju and Nadler (2013) note that self-efficacy dictates the manner in which an individual approaches a task, as well as influencing the goals that an individual will set as a desired outcome of the task. Both self-efficacy and goal setting inevitably affect the self-regulatory measures that an individual will employ throughout the completion of a task (Caraway *et al.* 2003). These include, but are not limited to, self-evaluation, self-monitoring and strategy use. Conversely, low self-efficacy could theoretically increase the likelihood of an individual employing negative self-regulation measures, such as self-handicapping, in order to protect self-portrayal or self-image (Urdan 2004). Due to a lack of belief in a positive outcome an individual can employ these measures and essentially guarantee their own failure before ever engaging with the task. Bandura (2012) provides a useful overview of self-efficacy within his broader social cognitive theory (Bandura 1986). Figure 1 provides a graphical representation of the manner in which self-efficacy can influence outcome expectations, goals and sociostructural factors. This "shows the paths of influence in the posited sociocognitive structural model of self-motivation and self-regulation of action. Self-efficacy is a focal determinant because it affects behavior both directly and by its influence on the other determinants" (Bandura 2012, p.14).

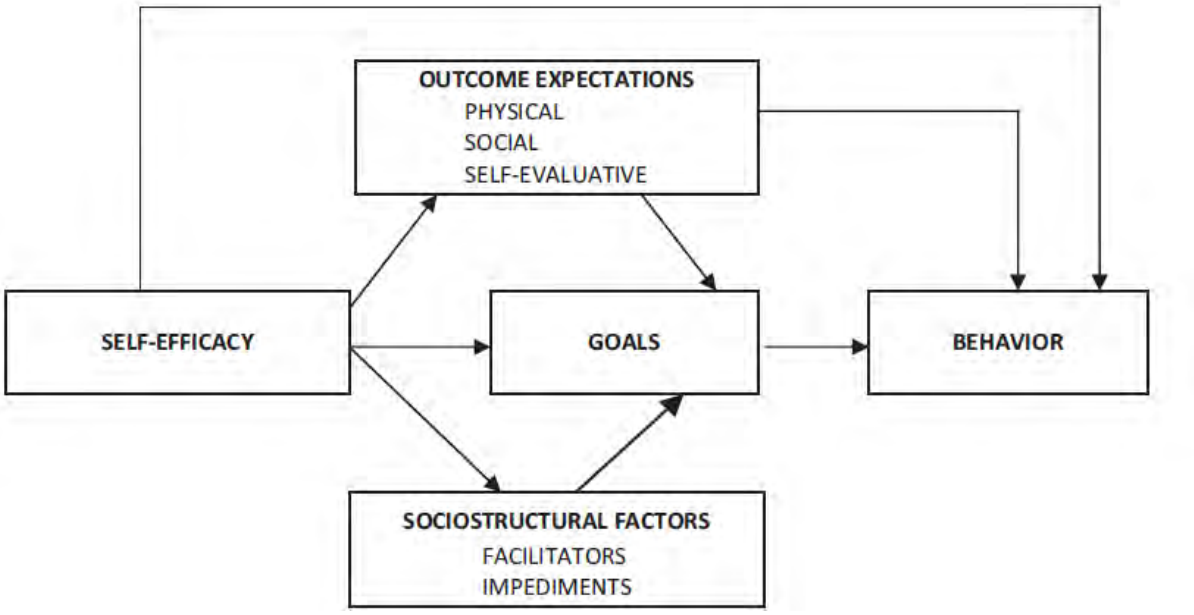

Figure 1. Self-efficacy, Goals and Behaviour from Bandura 2012, p.14

Typically studies examining self-efficacy impact tend to focus on one of the 'other determinants' (outcome expectations, goals or sociostructural factors) as seen in Figure 1. An example of one such study is that of Zimmerman *et al.* (1992) who conducted an experiment in which they utilised goal setting and self-efficacy in order to predict grades in a social studies class across a school term. They noted a 31% increase in prediction over prior grades. This study was later expanded in order to examine the Scholastic Aptitude test and a 35% increase in prediction was noted when examining final grades of a writing course (Zimmerman and Bandura 1994). Both of these studies supported the hypothesis that although prior grade attainment and aptitude tests are more commonly used predictors of future performance, when considered in conjunction with self-efficacy and goal setting, there is potential to significantly improve the desired predictive value. Zimmerman and Bandura (1994) also observed the predictive nature of self-efficacy in terms of university students personal standards. Their findings suggested that an individual's self-efficacy levels significantly predicted their personal standards, which they felt they had to meet in

order to consider their work to be of adequate quality, as well as goal setting. Heightened levels of self-efficacy were also shown to increase the use of learning strategies. A 15% to 18% shared variance was observed in a study of perceived verbal/mathematical efficacy and strategy use (Zimmerman and Martinez-Pons 1990). Overall self-efficacy has been found to account for 14% of variance in academic performance with a resultant effect size of .38 (Multon *et al.* 1991).

When discussing the relationship between self-efficacy and goal orientation Stevens and Gist (1997) noted that self-efficacy encourages the adoption of positive goal orientations within the same domain. This is supported by more recent studies examining this interaction within teacher training and performance (van Daal *et al.* 2014). This is also in line with the underpinning social cognitive theory (Bandura 1986), which suggests that self-efficacy can influence multiple facets of an individual's life including their goals (Caraway *et al.* 2003). As highlighted by Caraway *et al.* (2003), self-efficacy is enhanced by complimentary goal orientations. As a goal is achieved, self-efficacy is increased. This is supported by the theoretical foundations of self-efficacy theory, which highlights mastery experience as a primary source of self-efficacy (Zelenak 2015). A further study examined the predictive value of self-efficacy relative to goal adoptions (Liem *et al.* 2008). The authors observed a positive predictive value relative to mastery goal adoption and performance goal adoption. In addition there was a negative relationship between self-efficacy and performance avoidance goal adoption. Both goal orientation and self-efficacy have been found to be domain specific (Jiang *et al.* 2014). Jiang *et al.* (2014) also observed a link between goal orientation, self-efficacy development and environment. Ultimately Jiang *et al.* (2014) outlines the considerable similarities

between self-efficacy theory and goal orientation theory from a theoretical and empirical perspective.

Dearths and Disagreements

The following section outlines debates around current directions of self-efficacy research. The review focuses on four strands which are first examined independently, and later collectively, in order to identify potential dearths in the existing research base. The first strand examines the proposal of additional sources, the second examines the misinterpretation of sources, the third examines the importance of domain specificity when examining self-efficacy, while the fourth examines prior attempts to measure self-efficacy.

Proposed Additional Sources

Bandura (1986) originally theorised that individuals constructed their sense of self-efficacy by drawing from four sources of self-efficacy; Mastery Experience, Vicarious Experience, Social Persuasions and Physiological State. Over time additional sources were proposed. These were typically posited as domain specific, such as Palmer (2006) who in his examination of teacher self-efficacy suggested that content mastery could be interpreted as a unique source. This reflects a wider trend where new sources have been introduced as potential explanations for variance in results, as opposed to products of theoretically sound deductive research (Caprara *et al.* 2008). This, in turn, is often linked to poor instrument design, or adaptation, resulting in invalid results (Bandura 2012). According to Joo *et al.* (2013) and Bandura (2006) inferences drawn from these invalid results, including the proposals of additional sources, serve only to confound and degrade.

An example of these specialised additional sources is seen in Palmer's (2006) study, which examined the self-efficacy of science teachers and found that low self-efficacy could have negative affects on their future teaching endeavours. The author proposes additional sources of self-efficacy; cognitive content mastery, cognitive pedagogical mastery and simulated modelling. The results of the study suggest that cognitive pedagogical mastery was the major contributor to self-efficacy. However, this study's conception of domains and resultant subdivision of sources should be considered in light of later papers which raise concerns relative to domain specificity (Caprara *et al.* 2008, Bandura 2012). The proposal of additional sources which bear little theoretical or logical distinction from the existing sources (e.g. Palmer's (2006) cognitive content mastery and cognitive pedagogical mastery) has the potential to confound later discussions (Bandura 2012). Other researchers posit that a source can be considered independent if it is not directly linked to the domain of functioning (Schoon and Boone 1998).

Schoon and Boone (1998) had previously posited that content mastery acted as an additional and separate source. Schoon and Boone (1998) justify this as a distinct form of mastery due to its relationship with prerequisite knowledge as opposed to an executable task, in this scenario teaching. Rather they classify this as 'cognitive content mastery' and this exists as a separate source in their model. The authors argue that this distinction between domains of functioning adds credibility to this justification of a new and unique source. However, when considered in light of Bandura's (2012) paper this distinction could be interpreted as contradicting the underpinning theory due to the proposed additional source existing independently from the domain of functioning.

Misrepresentation of Sources

Previously outlined problems relating to proposals of additional sources are closely linked to misinterpretation of sources which sees researchers replace sources with unrelated or ill-conceived metrics (Bandura 2012). Williams and Williams (2010) outline how one of the more problematic, and widespread, of these misinterpretations sees the replacement of the mastery source with an objective performance metric. Acknowledging that one of the keymotivations behind Bandura's seminal work was the need for a response to a lack of acknowledgement of the role of cognition in the predominantly behaviourist work of the time; the misinterpretation of mastery is symptomatic of a wider misconception of theory in many research endeavours (Usher and Pajares 2009, Bandura 2012). Presenting a metric of performance as representing mastery experience fails to acknowledge the role of perception. For example, while some researchers directly equate previous performance metrics with mastery experience (Chin and Kameoka 2002) others request that participants verbally report previous grades (Matsui *et al.* 1990, Klassen 2004). Neither of these practices align with Bandura's theoretical underpinnings of self-efficacy theory and both fail to acknowledge the role of cognition when interpreting mastery source information. Expounding on the role of perception in mastery experience Bandura highlighted that "the same level of performance success may raise, leave unaffected, or lower perceived self-efficacy depending on how various personal and situational contributions are interpreted and weighted" (Bandura 1997, p.81).

Domain Specificity

An associated debate revolves around the explanatory and predictive value of self-efficacy relative to the domain in which it is being measured.

> "As self-efficacy theorists have pointed out, self-efficacy's explanatory and predictive power diminishes when these self-beliefs are assessed at broad levels of specificity and when they do not correspond faithfully with the outcome with which they are compared."
>
> (Usher and Pajares 2008, p.763)

When examining the sources of self-efficacy it is important that items or instruments situate the sources at an appropriate level of specificity. Assessing source information at too general a level will fail to provide reliable data regarding self-efficacy levels for a specific subject (Bandura 1997, Bandura 2006). This is especially important if comparing source information across multiple domains. As Bandura notes 'the weights assigned to different types of efficacy information may vary across different domains of functioning' (Bandura 1997, p. 114). In some studies, errors in specificity have been compounded by poor alignment between source items and overall self-efficacy measures.

> "The sources of self-efficacy also function best at appropriate levels of specificity, and when they correspond with the self-efficacy outcome they are designed to predict. It makes little sense, for example, to compare the sources of general academic self-efficacy with students' mathematics-specific efficacy judgments. Similarly, assessing the sources at too general a domain level would offer little help in predicting students' subject-specific self-efficacy."
>
> (Usher and Pajares 2008, p.763)

These issues can be seen in studies that do not employ enough items and as a result fail to acknowledge the complexity of the sources (Panagos and DuBois 1999). Even more worrying is the widespread practice of not reporting critical data relating to newly developed scales (Özyürek 2005, Stevens *et al.* 2006, Bates and Khasawneh 2007). This makes it difficult, if not impossible, to ascertain whether theoretical guidelines have been adhered to (Usher and Pajares 2008). This, in turn, highlights

the need for research that clearly defines the specificity of the domain and examines the sources accordingly if our understanding of self-efficacy formation is to be further developed. This is representative of a wider debate within the field that includes two of the most heavily contested aspects of self-efficacy theory, domain specificity and source influence. This is discussed by Bandura (2006) who cautioned that researchers involved in the investigation of additional sources need to be cognisant of the domain of functioning and theoretical distinction of sources. This debate is focused on the role that each of these factors plays in the development of self-efficacy (Usher and Pajares 2008, Joo *et al.* 2013, Honicke and Broadbent 2016). The previous studies highlight the need for researchers to be cognisant of the influence of the domain when examining self-efficacy. This is especially critical when selecting or creating items designed to measure self-efficacy (Bandura 2006, Pajares 2007). In order to examine source influence, the current study selected a task that presented a uniquely defined domain in the form of a computerised maze navigation task. This task is especially suitable as it requires relatively little prerequisite skills or knowledge while also presenting a platform that can track longitudinal data such as engagement and performance.

Measurement of Self-efficacy

Considering the debates outlined previously regarding the definition of sources, and the nature of self-efficacy itself, there exists a predictable lack of consensus when it comes to measuring the sources of self-efficacy (Bandura 2012). A considerable amount of studies that have examined the sources of self-efficacy have used adapted versions of the Sources of Mathematics Self-Efficacy Scale (SMES) originally developed by Lent *et al.* (1991). While originally developed for use in the domain of mathematics subsequent studies have adapted the scale for use in broader academic

and social settings (Anderson and Betz 2001, Smith 2001, Britner and Pajares 2006, Usher and Pajares 2006). Alternatives include Hampton's (1998) Sources of Academic Self-Efficacy Scale that was later the subject of considerable validity orientated scrutiny by Hampton and Mason (2003). The common links between the scales mentioned above include considerable research aimed at developing and validating the instruments in addition to a sound interpretation of theory in their conception. However, each study primarily focused on a unidirectional relationship of self-efficacy on performance. The studies mentioned above did not facilitate an examination of how the activity, or task, influenced self-efficacy. This results in studies that focus on variance of outcome scores that can be attributed to self-efficacy. The lack of research examining the impact of the task on self-efficacy is noted and identified as a potential theoretical void that this study has the potential to inform.

Examples of a less robust manner of scale development, or reporting, can be seen in the work of Bates and Khasawneh (2007) or Stevens *et al.* (2006), both of which failed to report source items. A compounding problem is the use of alternative measures, such as performance metrics discussed in terms of mastery experience previously (Chin and Kameoka 2002), to represent one or more of the sources. Misrepresentation of sources extends beyond the common ill-advised substitution of mastery experience with performance measures. The vicarious experience source has proven to be the most elusive source in terms of reliability and replicability (Usher and Pajares 2008). As highlighted by Harris (1995) not only do peers and adults differ in their impact as vicarious sources, but these disparate sources differ further in their influence as an individual develops. This supports the practice of including source items examining peer and adult influence on the vicarious source as

seen in the widely adapted scale originally developed by Lent *et al.* (1991). Usher and Pajares (2009) support the inclusion of items examining peer and adult influence on the vicarious source suggesting that scales that fail to do so offer an incomplete perspective. They also suggest that even scales that employ items that allow for a more complete examination of the vicarious source are likely to return poor reliability values due to the disparate influences drawn from peers and adults. This is broadly supported with the majority of studies that have directly examined this source reporting poor reliability coefficients (Smith 2001, Usher and Pajares 2006, Zelenak 2015). Similarly when assessing social persuasions there exist numerous examples of misconceptions, and practices, that do not reflect social persuasions as theorised by Bandura (2012). For example Hampton (1998) examined the nature of instructions received by participants while others have sought reports on expectations of future performance (Chin and Kameoka 2002). In contrast studies that ask participants to report the level of encouragement typically align with theory and also report relatively high reliability values (Lent *et al.* 1996, Usher and Pajares 2009).

When examining the physiological source researchers have typically examined the negative influence of this source drawing heavily from existing scales such as Lent's (1991) adaptation of the Betz (1978) revised Math Anxiety Scale. Similar anxiety or negative arousal items have been adapted by a considerable number of researchers working in the area (van Dinther *et al.* 2011, Joo *et al.* 2013, Jiang *et al.* 2014, Lee 2015, Zelenak 2015, Honicke and Broadbent 2016). While studies that focus on negative aspects of physiological arousal have reported consistently high reliability values (Usher and Pajares 2009), they fail to acknowledge the positive dimensions of the physiological source (Bandura 2012). In summary disagreements relating to

measuring the sources of self-efficacy can be typically categorised as misrepresentations, such as using past performance metrics to represent mastery experience, or misinterpretations of theory as seen in poor item or methodological design (Bandura 1986, Bandura 1997).

> "There is no all-purpose measure of perceived self-efficacy. The "one measure fits all" approach usually has limited explanatory and predictive value because most of the items in an all-purpose test may have little or no relevance to the domain of functioning."
>
> (Bandura 2006, p.307)

This highlights the need for cognisance of the domain and theoretical underpinnings when utilising a pre-existing scale. Bandura (2006) continues his criticism of uninformed adaptations when he warns that alterations, or scale construction, must focus on the domain of functioning that is to be examined. Failure to do so will result in increased results indicating domain interplay. Although interaction between domains, when similar sub skills are employed, is acknowledged; poor design will lead to invalid inflated results of this interaction (Bandura 2006, 2012). This highlights the problematic aspects of exploring the relationship between self-efficacy and performance in a broad domain. The current study has the potential to contribute to this research area by employing a task that is more defined and less dependent on prerequisite skills in order to allow for a clearer examination of this relationship.

Cognisant of the points raised previously, the mechanism of self-efficacy measurement used in the later stage of the current study had to satisfy the following criteria:

1. Self-efficacy must be measured as a product of sources rather than a composite.

2. The scale must attempt to measure all four sources in a manner that aligns with the tenets of self-efficacy theory.
3. Necessary alterations to the scale should be minimal while also further focusing the domain of functioning.

A review of existing scales resulted in the selection of the "Sources of self-efficacy in mathematics" scale (Usher and Pajares 2009). This scale underwent a considerable development and review process while also satisfying all three criteria listed above. The necessary adaptations were minimal and did not detract from the "relevance to the domain of functioning" as cautioned by Bandura (2006, p.307). A notable strength of the selected scale was the considerable lengths the authors went to in order to examine evidence of construct validity.

Construct Validity

Construct validity has proved to be one of the most vehemently debated aspects of self-efficacy in recent years (Multon *et al.* 1991, Anderson and Betz 2001, Schunk and Pajares 2009, Bandura 2012). Perhaps one of the most pressing concerns stems from a widespread misunderstanding of the fundamental aspects of self-efficacy theory.

> "All too often, this belief system is treated as though it is a generalized trait. In fact, people differ in their efficacy, not only across different domains of functioning but even across various facets within an activity domain. Consequently, there is no single all-purpose measure of self-efficacy with a single validity coefficient. The construction of valid self-efficacy scales requires sound conceptual specification of the determinants governing performance in a given domain of functioning and the impediments to realizing desired attainments."
>
> (Bandura 2012, p.15)

Bandura (2006) describes self-efficacy as being concerned with individuals' and their ability to achieve favourable outcomes. Using this description he highlights that self-efficacy should not be considered a generalisable trait. This highlights individuals' unique self-efficacy beliefs which can vary not only across domains but even across various aspects of that domain. For example, an investigation of Math self-efficacy as a generalisable trait would assume that an individual would maintain a consistent self-efficacy level when encountering various mathematics based tasks. The reality of the complex nature of self-efficacy could allow a person to exhibit high self-efficacy when encountering general arithmetic but also demonstrate low levels when encountering a specific branch such as long division (Lent *et al.* 1991, Lopez and Lent 1992). Bandura (2006, p.15) acknowledges the intricacies of belief systems in the design of scales and stresses that as a result of these intricacies there can be no universal measure of self-efficacy. Many of the issues discussed here arise from a fundamental misunderstanding of self-efficacy sources as alluded to by Bandura (2012). A tendency of researchers to omit information regarding construct validity combined with the consistently low reliability scores of vicarious experience support Usher and Pajares (2008) assertion that many measures used to address this source have been inadequate.

> "Validity is an overall evaluative judgement of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of interpretations and actions on the basis of test scores or other modes of assessments."
>
> (Messick 1995b, p.741)

Messick's restructuring and unifying of validity was the culmination of a continuous effort of the wider community of practice dating back to individualised conceptions of validity broadly employed in the 1940's (Messick 1995a). In the creation of his

Unified Theory of Construct Validity Messick identified 6 aspects of construct validity. These 6 aspects: Consequential Content, Substantive, Structural, External and Generalisable will be used as a framework in order to examine validity related implications and limitations of the current research project.

The consequential aspect refers to negative outcomes if the scores, or measurements, are invalid. This provides an interesting point from which to examine the manipulation of scores within this study. A key premise of the current study is the manipulation of the practice task in order to impact participants' perceived self-efficacy. In this manner participants' engagement rates, and as a result final performance, can be reduced by negatively influencing their self-efficacy for the given task. To ensure that hidden difficulty variations were responsible for this influence multiple rounds of testing were employed that examined the influence of additional variables. This ensured that "low scores [did] not occur because the assessment is missing something relevant to the focal construct that, if present, would have permitted the affected persons to display their competence" (Messick 1995b, p.746). Rather low scores attributed to manipulation of the construct in question aligned with theoretically predicted outcomes suggesting a valid manipulation of the construct. This is further supported by the manipulation of the practice task but not the final retest version of the task. The comparatively equal variances attributed to high and low difficulty with respect to the overall round means suggest that low scores were not a result of "the measurement contain[ing] something irrelevant that interferes with the affected persons' demonstration of competence" (Messick 1995b, p.746).

Content relevance is concerned with defining the limits of the construct to be defined. This requires the establishment of attitudes, knowledge, motives, skills and

other attributes to be examined by the assessment. It should be noted that the term assessment in this context is defined as an activity including tests, inventories, scales etc. This can be achieved by analysing the task itself, in terms of a performance assessment, or through the use of domain theory. Domain theory is defined as:

> "[…]scientific inquiry into the nature of the domain processes and the ways in which they combine to produce effects or outcomes. A major goal of domain theory is to understand the construct-relevant sources of task difficulty, which then serves as a guide to the rational development and scoring of performance tasks and other assessment formats."
>
> (Messick 1995b)

This aspect of validity is especially relevant to the current study as it addressed two key concepts that are fundamental to the study; domain and difficulty of a task. The concept of a defined domain is a key component of self-efficacy theory and critical to its explanatory and predictive power (Bandura 1977, 2006). Self-efficacy theory provides a thorough theoretical conceptualisation of the associated attitudes, knowledge, motives, skills and other attributes when completing a task within a domain (Schunk and Pajares 2009) and highlights the importance of the researcher being cognisant of the domain in any research endeavour that seeks to examine self-efficacy (Usher and Pajares 2008). This study intentionally selected a task that allowed for a comparatively well-defined domain in order to observe the influences of varying difficulty.

Within this defined domain this study intentionally manipulates the difficulty of the task which speaks to the clearly linked construct-relevant sources of task difficulty above. The resultant disparate outcomes, when compared across difficulty groups, suggest that the source of difficulty was correctly identified and subsequently manipulated. Traditionally this aspect of validity would have concerned itself with

the likes of mathematical ability assessments where multiple levels of difficulty would exist within an assessment in order to enhance discriminant validity. The task in this study employs a novel twist in this regard where difficulty is manipulated unbeknownst to the participant. Inferences of discriminant validity can instead be inferred from differences in outcomes based on the categorical difficulty settings employed. This is also supported by the additional self-efficacy source data which aligns with the performance based data to provide a unique perspective relative to the domain processes and outcomes.

The substantive aspect of validity revolves around the role of theory in identifying the processes to be illuminated through the employment of the assessment or task. This is typically examined in the manner which the tasks assess a suitable spread of domain processes and domain content. This is a further reminder that this type of validity is typically used in examining much broader domains with a considerable breadth of content. In the context of this study content is not the primary focus. Rather the processes in terms of task difficulty impact on outcomes and self-efficacy formation are central. There exists considerable theoretical guidance in this regard and the process relating to self-efficacy formation and performance outcomes have been previously addressed and interpreted in the context of theory (Bandura 2012).

The structural aspects of construct validity require the test to be rationally consistent with the theoretically predicted outcomes of the construct so that these outcomes can be measured in such a way as to be considered valid (Messick 1994). The current study identified behavioural manifestations indicative of the construct in question and recorded data in the pre-test, post-test and longitudinal data from the practice period in order compare these manifestations with theoretical predictions. In order to further examine these relationships each round was manipulated in a stratified

manner and data was then compared internally in order to examine whether the theoretically informed manipulation resulted in behavioural manifestations indicative of a rational interpretation of the construct. The additional perspectives of different rounds allowed for an examination of the consistency of this influence as well as an indication of potential unidentified variable influence.

The external aspect of construct validity relates to expected high and low behavioural manifestations, and non-assessment behaviour, of the construct in question. Noting that a key premise of the current study is the manipulation of task properties, and by extension manipulating the construct in question, the resulting high, medium and low difficulty groups allow for adequate comparisons of behaviour. In addition, researchers should also be concerned with "interactive relations implicit in the theory of the construct being assessed" (Messick 1995b, p.746). Examples of design and analysis steps taken to examine various interactive relations can be seen in the comparison of round 1 and round 2 data in order to examine the role of environment. Similar steps were taken in the design and comparison of round 3, 4 and 5 in order to examine a suspected interaction effect observed in round 3 data. From a broader perspective this aspect can be said to entail convergent, discriminant and predictive qualities. The comparisons of multiple rounds of testing and the additional scale data from round 5 suggest convergent validity. The examination of discriminant validity is less direct due to ubiquitous nature of self-efficacy in the most popular theories of self-regulation. As noted by Sitzmann and Ely (2011) the construct of self-efficacy is present in six out of seven of the most widely examined and popular theories. To compound the issue further self-efficacy theory is said to incorporate elements of many of the most broadly accepted self-regulatory theories. The issue of discriminant validity has proven to be

problematic and this is further evident in the level of correlations between constructs noted in multiple large scale reviews (Linnenbrink and Pintrich 2002, Usher and Pajares 2008, Hulleman *et al.* 2010, Burnette *et al.* 2013, Honicke and Broadbent 2016).

The concern of the generalisability aspect of construct validity, in its simplest form, can be encapsulated by the question of whether the results can be generalised across disparate groups, settings and tasks. Noting the 5 rounds of testing employed in this study initially suggests considerable grounds for generalisability, however, it should be noted that all participants were of a similar age and recruited from a single school. A similar study, albeit with a different focus, noted patterns consistent with those observed in this study (Lynch *et al.* 2013). The comparison of round 1 and round 2 allows for a limited discussion relating to setting but once again the participants were drawn from the same school population. However, it should also be noted that the population of the school represents an especially diverse group of students in terms of national aptitude test scores.

The previous six aspects of construct validity should not be considered in isolation. Rather they provide a useful framework with which various questions relating to construct validity, in the context of the current study, can be examined. This examination takes the form of theoretical rationales and deductive argumentation. “[T]he relation between the evidence and the inferences drawn that should determine the validation focus" (Messick 1989, p.16). The inferences drawn from the data gathered in this study were refined as further competing rationales for observed variances were discounted using additional rounds. Early data was viewed as behavioural manifestations indicative of construct manipulation. Comparisons between rounds supported the influence of the manipulated difficulty variable and

behaviour aligned with theoretical predictions. Subsequent data further supported the impact of this variable in alternative settings and as having a considerably greater impact than the reward variable. Further data from the sources of self-efficacy scale (Usher and Pajares 2009) used in round 5 support the inferences drawn from earlier behavioural manifestations of self-efficacy impact. Previous inferences hinged on differences in behaviour aligning with altered self-efficacy levels. The scale data provides further support with markedly different self-efficacy source values noted between difficulty groups. The combination of theoretically linked behavioural manifestations of construct manipulation in conjunction with the striated self-efficacy source values suggest convergent validity.

Potential Gaps in the Literature

Self-efficacy as a research topic has seen exponential growth since Banduras seminal paper (Schunk and Pajares 2009). This has resulted in a wealth of unique perspectives, but has also led to problematic research designs that some renowned researchers believe are detrimental to the continuing development of the theory (Pajares 1997, Bandura 2012, Sitzmann and Yeo 2013). Criticisms relating to design focus on two key points; misalignment with the theoretical underpinnings of self-efficacy theory and insufficient methodological rigour resulting in data that fails to address a lack of empirical testing within self-efficacy theory (Tsai *et al.* 2011, Bandura 2012, Zelenak 2015).

Misalignment with the underpinning theory has resulted in problems relating to scale design and source relativity (Bandura 2006). As highlighted by Bandura "The construction of valid self-efficacy scales requires sound conceptual specification of the determinants governing performance in a given domain of functioning and the

impediments to realizing desired attainments" (Bandura 2012, p. 15). Many of the research projects discussed in this review examine performance relative to self-efficacy. The majority of these only examine the impact of self-efficacy on performance. The scarcity of empirical evidence examining the posited bi-directional nature of performance and self-efficacy is problematic as it fails to explore the influence of task properties on self-efficacy formation (Williams and Williams 2010). Further evidence of widespread misconceptions relating to the role of the task in the formation of self-efficacy is seen in the equation of performance figures (grades) with mastery experience (Bandura 2012). This fails to acknowledge the role of perception and cognition in the processing of self-efficacy source information. Furthermore it has led to data which is not useful when examining the sources, and development, of self-efficacy (Honicke and Broadbent 2016).

This presents a considerable gap in the literature highlighted by this review. An examination of the influence of perceived performance on self-efficacy in a manner that is consistent with the theoretical underpinnings of self-efficacy theory is necessary if the hypothesised reciprocal nature of performance and self-efficacy is to be supported. This requires a defined domain of functioning and an understanding of what constitutes success in the selected task. In addition the chosen task must be suitable for manipulation. "Perceived efficacy should be measured against levels of task demands that represent gradations of challenges or impediments to successful performance" (Bandura 2006, p.311). Linked to this point is the need for an examination of the sources of self-efficacy that is true to the theoretical underpinnings. In an examination of the impact of performance on self-efficacy an examination of the sources of self-efficacy will provide more valuable data than a composite measurement (Usher 2009).

Misinterpretation of theory compounded by poor methodological design sees a proliferation of papers that treat self-efficacy as a composite construct (Usher and Pajares 2008). The design of scales that fail to examine the role of each source has limited value for the further development of theory (Bandura 2006). Noting that source dominance and influence are currently some of the most contested elements of self-efficacy theory; this practice is especially detrimental and highlights the need for future research to be cognisant of rigorous methodological design. An associated dearth in the literature is a logical consequence of a scarcity of research examining the sources of self-efficacy; the examination of factors which influence the development of the four hypothesised sources of self-efficacy. Recent research highlights the lack of empirical data relating to this area and call for future research to address this deficiency (Bandura 2012, Parker *et al.* 2014, Zelenak 2015).

Research examining the outlined dearths and ongoing debates within self-efficacy theory is vital. As further interest in the theory grows, the potential for further theoretical refinement is dependent on research that is true to the tenets of self-efficacy theory. In addition to furthering the theoretical evolution of self-efficacy, the resolution of the points raised will lead to more effective real world impacts from studies that seek to utilise self-efficacy in order to increase desirable behaviour altering outcomes. Previous studies have highlighted the positive impacts of self-efficacy in the domains of health (French *et al.* 2014), education (Komarraju and Nadler 2013) and technology use (Chiou and Wan 2007). As the benefits of increased self-efficacy become ever more apparent there has been a rise in intervention style studies that specifically aim to enhance self-efficacy in order to enhance desired performance or behaviour (Schunk and Ertmer 2000). This project has the potential to contribute to a discussion around the factors that enhance self-

efficacy but also to the nature of the relationship between performance and self-efficacy.

CHAPTER THREE: METHODOLOGY

> "Perceived efficacy should be measured against levels of task demands that represent gradations of challenges or impediments to successful performance. Self-efficacy appraisals reflect the level of difficulty individuals believe they can surmount. If there are no obstacles to overcome, the activity is easily performable and everyone is highly efficacious."
>
> (Bandura 2006, p.311)

Introduction

The following chapter outlines the broader context of the study. The purpose, experimental design, participant profile and sampling method are examined in order to explicitly outline the influence of these factors on the research project.

Context

The school involved in this study is situated in the south of Ireland and has been designated as being disadvantaged and therefore is part of the Delivering Equality of Opportunity in Schools (DEIS) initiative (Department of Education & Science 2005). It is a mixed-sex school with a 68% male contingent in the first year enrolment in the school year 2012-2013 (Rounds 1 and 2), 62% male contingent in the school year 2013-2014 (Rounds 3 and 4) and a 65% male contingent in the school year 2014-2015 (Round 5). The school in question accommodates a considerable range of ability, with students scoring in both the lowest and highest 10% in national standardised test scores.

Purpose

The main aim of the current study is to examine the relationship between task difficulty, engagement, performance and self-efficacy formation. The impact of task difficulty on these outcomes is examined within a theoretical framework of self-

efficacy. By examining these outcomes the current study aims to provide empirical data that will further inform two prominent ongoing debates within self-efficacy theory. Firstly the impact of task performance on self-efficacy formation; secondly the hypothesised reciprocal nature of self-efficacy and performance.

The majority of studies listed previously cannot provide data that facilitates examination of this reciprocal relationship, as proposed by the existing theoretical base (See Figure 7), due to their cross sectional designs. This is partly due to the complexity of the domains in which self-efficacy was examined. A more defined and constrained domain, such as the one employed in the current study, presents an opportunity for manipulation and observation that has the potential to provide insight into the relationship between task difficulty, engagement, performance and self-efficacy formation.

Research Questions

What is the relationship between task difficulty and engagement in the given task?

What is the associated relationship between task difficulty and performance in the given task?

What is the relationship between task difficulty, engagement, performance and self-efficacy formation in the given task?

Experimental Design

In order for a design to be considered a 'true experiment' it must include manipulation, control, random assignment and random selection (Cohen *et al.* 2000). Manipulation refers to a purposeful alteration of a variable by the researcher. Control refers to a fraction of participants who are not subject to the manipulated variable.

Random assignment refers to the manner in which participants are assigned to groups within the study while random selection refers to the manner in which participants were selected from the wider population to take part in the study. These elements in addition to systematic procedures are employed to reduce error or bias. While the current design employs manipulation, control, and random assignment it can claim limited random selection as the available population was restricted by the school setting. Noting this limitation in addition to the setting of the experiment, and associated unidentifiable variables, the current design is positioned more completely within a quasi-experimental definition (Campbell and Stanley 2015).

Manipulation

In rounds 1, 2 and 5 the difficulty of the task was manipulated in order to create a low difficulty, medium difficulty and high difficulty version of the task. The difficulty variations were achieved by manipulating the speed of antagonistic characters within the task. The high difficulty version of the task had characters that moved 10% faster that the medium difficulty task. Conversely characters from the low difficulty version of the task moved 10% slower than the medium difficulty task. Difficulty manipulation based on speed variation is supported by Gilbert *et al.* (2012). Pilot testing prior to the commencement of the study led to the development of difficulty variations that were not evident to the user.

Round 3 employed an additional manipulation in the form of the points awarded for each positive action within the task. This led to a proportional reward system where the high difficulty task was paired with a high reward system. This high reward system was achieved by awarding 10% more points for each positive action within the task. Conversely, the low difficulty task was paired with a reward system where the participant received 10% fewer points for each positive action. The medium

difficulty task was not altered. Round 4 maintained the reward system manipulations used in round 3 but removed the difficulty manipulation. This resulted in a high reward, medium reward and low reward version of the task.

Control

As discussed in the previous section each round consisted of three groups defined by the level of manipulation of the variable in question. The main control within the study is in the form of the medium group within each round. This control design should be considered in light of warnings relating to the positive impact of control groups being limited by the amount of participants involved (Festinger *et al.* 1970). In order to address concerns relating to this limitation between group within round analyses were carried out and are discussed in detail in the later results chapter.

Random Assignment

Participants are randomly assigned to one of three groups within each round. As previously outlined the groups are defined by the task settings resulting in a high, medium and low group. The random assignment process takes place directly after the pre-test where the participant selects a USB device from a container. Earlier applications of the software (Lynch *et al.* 2013) suggest that alterations to each version were not identifiable to participants. Furthermore, the USB devices bear no identifying marks relating to the altered variable.

Random Selection

Participants were selected from a first year population of a second level co-educational school in the south of Ireland. A list of all first year students was obtained in order to create a sampling frame (Hopkins *et al.* 1987) and each student was assigned a randomised non-repeating number. The list was then arranged in

descending order based on the randomly assigned number. The required number of participants for the relative round of testing was then selected. Although the selection of participants was randomised; the selection of the school was not. The selection of the school was based on a convenience sample due to restrictive timeframes and logistical constraints.

Validity

Internal validity reflects the degree to which a study minimises systematic error. In order for inferences to be internally valid, in a quasi-experimental design, a causal relationship must be demonstrated between an independent and dependent variable (Brewer 2000). Shadish *et al.* (2002) suggest that an inference of causality can be considered valid if it meets all of the following criteria: Co-Variation, Non-Spuriousness and Temporal Precedence. Co-variation refers to the nature of the relationship between the cause and effect i.e. the observed effect must be related. Temporal precedence reflects the time and sequential nature of cause and effect i.e. the cause must occur before the effect. The final criterion, non-spuriousness, requires no alternative plausible explanation for the co-variance outlined previously. The design of this study was informed by the three preceding criteria. This is reflected in the design of each round but also in the sequential nature the rounds. The primary variable of interest, task difficulty was examined in the context of other potential confounding variables in an effort to strengthen inferences of causality. Liebert (1995) suggests that when observed changes, or comparisons, of the independent variable can be confidently linked to the dependent variable and alternative explanations can be discounted then causality can be claimed to be internally valid.

Although quasi-experimental designs typically claim considerable internal validity it is essential that the researcher is cognitive of the implications for external validity

when designing a study (Creswell 2013). External validity is defined by the extent to which inferences drawn from a study can be applied in other contexts and to other people (Mitchell and Jolley 2012). Threats to external validity can be described as any undesired interaction with the independent variable. Dipboye and Flanagan (1979) examine what they call the basic dilemma of the social psychologist. This dilemma comes from the view that efforts to increase external validity often leads to negative impacts on internal validity with the converse also being true. Exercising control to reduce the impact of unidentified variables can result in a sterile lab style experiment. As a result, the generalisability of the findings is limited due to the experimental environment bearing little resemblance to the normal operating environment of the participant. Field experiments are proposed by some to enhance generalisability however one should not assume inherent external validity (Liebert 1995, Shadish *et al.* 2002). The participants of this study completed the task throughout a typical week in an attempt to increase external validity. While the rationale for the selection of the task employed within this study has been previously explored, it is important to note that generalisability to more complex tasks is not assumed.

While debates around validity continue there is a broad consensus that replication is the truest measure of generalisability (Neuman 2005, Onwuegbuzie and Leech 2005, Weber *et al.* 2011). The sequential nature of the current study allows for comparisons between rounds with different populations and an examination of whether the impact of difficulty is replicable. However, this is limited to the specific school within which the current project took place.

Data Collection

As previously outlined the current study consisted of five rounds. Each round examined the influence of altered task settings, including difficulty and reward, on engagement and performance. Each round utilised a computerised maze navigation task and identical testing procedure. Participants completed the pre and post-test in groups of six to eight using one laptop. The same laptop was used by all participants in the pre and post-test procedure. The original version of this task was developed for use in a similar study that examined the impact of task difficulty in an alternative school setting (Lynch *et al.* 2013). Three versions of the practice task were created for each round of testing. These versions were based on altering the task variable in question in a positive and negative manner. This resulted in a High, Medium and Low version of the task for each round defined by the variable that was manipulated (Difficulty and/or Reward).

All participants were selected from the first year cohort of a co-educational school in southern Ireland. Participants for rounds one and two were randomly selected from students completing their first year of second level education[3] in the school year 2012-2013. Participants for rounds three and four were randomly selected from students completing their first year of second level education in the school year 2013-2014 while round five participants were selected from students completing their initial year of second level education in the school year 2014-2015. Participants were selected for invitation to take part in the study using a random number generator. Once every student in the school year had been assigned a random non-repeated number the student list was sorted to form a list based on the value of the randomly assigned number. The version of the task that each participant received,

[3] Irish second level education corresponds to middle school American designation (Buchanan and Fox, 2008). Students are typically 12 years of age upon entry.

forming three groups within each round of testing, was also randomly assigned. Participants selected their USB device, with one of three versions of the task uploaded to the device, once they had completed the pre-test.

Commonalities

Each round of testing employed the computerised maze navigation task described previously. All rounds employed a test/retest approach. This required all participants to complete the task on standardised settings before and after the practice period. This allowed for each individual's initial aptitude and overall improvement to be assessed. Once a participant completed the pre-test they were issued with a practice version of the computerised task on a USB drive. This could then be used to practice the task over the following five day period before the participant completed the standard post-test version of the task. The practice version that each participant received was chosen at random from one of three versions. The three versions of the task are derived by altering the variable in question resulting in a high version, medium version and low version. The participants were not informed of the hidden task variations. Throughout the practice period data relating to the task practice attempts is stored on the USB device for later analysis. This includes time, date and duration of each practice attempt. In addition data relating to performance within the task was also recorded.

Participants

Round 1 (n=62) and round 2 (n=62) were recruited from the 2012-2013 1st year school population. Round 3 (n=61) round 4 (n=66) were recruited from the 2013-2014 1st year population while round 5 (n=66) was recruited from the 2014-2015 1st

year population. The mean age of each first year cohort at time of enrolment was 12 years of age.

Materials

Information sheets and associated consent documents were distributed one week prior to the start of the data gathering. These documents are presented in Appendix C (p. 170). The opt-in nature of the research design required participants to return the consent section of the form in order to take part in the research project. All consent forms and associated material were approved by the appropriate ethics board prior to use.

The sources of self-efficacy scale employed in this study is a modified version of Usher and Pajares (2009) scale originally designed for use in the domain of mathematics. The original version is included with the accompanying altered version in Appendix B (p. 172). The scale and implications of alterations are examined in detail in a previous section (See p. 36). Evaluations of the altered version, in terms of reliability, are presented in a later section (See p.102).

The computerised maze navigation task employed as part of this research was derived from a non-deterministic version of the game[4] in order to control for learned patterns of character movement. Pac-Man was chosen as it requires comparatively few prerequisite skills or knowledge when compared to a mathematical or education based task. An adaptable version of the Pac-Man computer game was developed which included an automatic performance logging system. This logging system automatically recorded participants' performance data at 30-second intervals throughout the task and also at the conclusion of the task. Participants were presented with one of three versions of the practice task in each round. An outline of the manner in which versions were manipulated per round are presented in a previous section (See p.53). The interface of the task is presented in Figure 2.

[4] The altered game versions employed were originally derived from a non-deterministic version of the popular Pac-Man game. The non-deterministic properties of the game characters prevented participants from learning movement patterns which could have altered participant's performance. The difficulty variations and data logging systems were previously developed for use in a similar study by Lynch et al. (2013).

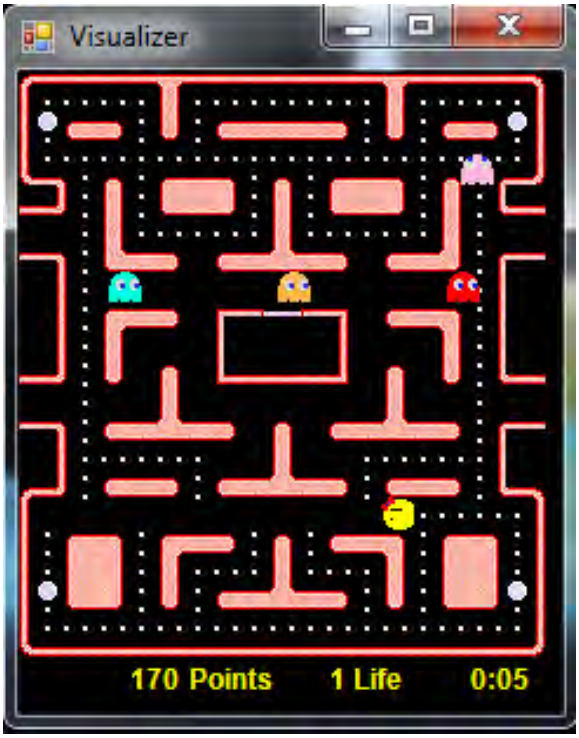

Figure 2. Task Interface

Melissinos and O'Rourke (2012) offer a complete history and discussion outlining the many iterations of this pervasive game.

Data Collection

The following section outlines the structure of each round including the influence from the previous rounds. A complete outline of the differences and sequential logic of the round design is provided at the end of this section. All rounds used a pre and post-test design. Each participant completed the pre-test on a standard version of the computerised maze navigation task in order to gauge initial performance. This pre-test took place in a spate room, participants were taken in groups of 6-8 from class and completed the task on identical laptop computers. Once the pre-test was complete participants were randomly assigned one of the three practice versions of the task. The task was contained within a USB drive. Participants were then free to practice with their randomly assigned modified practice version of the task for the

prescribed practice period. Participants were unaware of the alterations made to the three practice versions. Finally they completed the post-test, again using the standard version, in order to gauge improvement. USB drives were then returned containing data relating to the practice period. The manner in which the practice task was modified is outlined for each round below.

Round 1(Difficulty Variable Out-of-School)

Round 1 examined the influence of task difficulty in an out of school environment. The three practice versions of the task were derived from the original version by altering the difficulty of the task resulting in a High Difficulty, Medium Difficulty and Low Difficulty. Difficulty was altered based on the speed of the antagonistic characters within the task (Gilbert *et al.* 2012). An earlier study that originally developed the task in question noted that alterations were not identifiable in their application (Lynch *et al.* 2013). The practice period was completed over a five-day period when the participants were not attending school. This allowed for an examination of the impact of varying task difficulty on engagement and performance in an out of school environment.

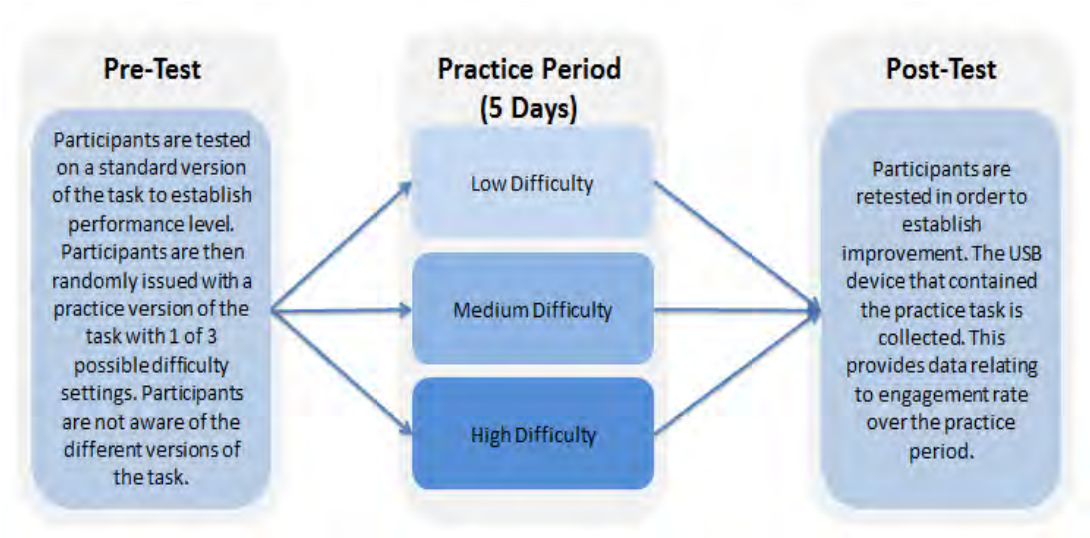

Figure 3. Round 1&2 Testing Structure

Round 2 (Difficulty Variable In-School)

Round 2 used the same task settings described in the previous round resulting in a High Difficulty, Medium Difficulty and Low Difficulty version of the practice task. In this round the practice period took place again over the course of a standard school week (5 days). This allows for the examination of whether the influence of varying difficulty is consistent across different practice environments. All subsequent rounds employed a practice period that took place over a typical school week in order to examine the impact of potentially confounding variables.

Round 3 (Proportional Reward In-School)

Round 3 examined the impact of varying reward in addition to varying difficulty levels. The task was further altered from the previous round resulting in a High Difficulty/High Reward, Medium Difficulty/Medium Reward and Low Difficulty/Low Reward. The reward level was altered by adjusting the amount of score points received for each positive action within the task. The percentage alteration of reward was proportional to the percentage increase in antagonist's character speed used to create the earlier difficulty variations. This is referred to as a proportional reward system in the later discussions. In this manner, the impact of varying reward on engagement, relative to the negative impact of difficulty observed in the previous rounds, could be explored.

Figure 4. Round 3 Testing Structure

Round 4 (Reward Variable In-School)

Round 4 used the reward variations used in the previous round but did not vary difficulty level. This resulted in a High Reward, Medium Reward and Low Reward version of the task. This allowed for an examination of the impact of an altered reward system without the influence of the varying difficulty levels observed in previous rounds.

Figure 5. Round 4 Testing Structure

Round 5 (Difficulty Variable with Sources of Self-Efficacy Scale In-School)

Data from rounds 1 to 4 supported a hypothesised influence of task difficulty on self-efficacy formation. Results relating to overall improvement and engagement aligned with theoretically predicted outcomes suggesting varied hidden task difficulty levels impacted self-efficacy. While the previously collected data aligned with theoretical predictions, the manner in which task difficulty was influencing self-efficacy still remained unclear. In order to examine this influence, participants also completed the adapted Sources of Self-Efficacy scale (Usher and Pajares 2009) at pre and post-test. This scale was originally designed for use within the domain of mathematics and the wording was modified for use with the maze navigation task. In line with Bandura's guide for constructing self-efficacy scales (Bandura 2006) and with the broader theoretical underpinnings of Self-Efficacy theory (Bandura 2012) it was anticipated that the increased specificity of the domain would have positive effects on the validity of the results.

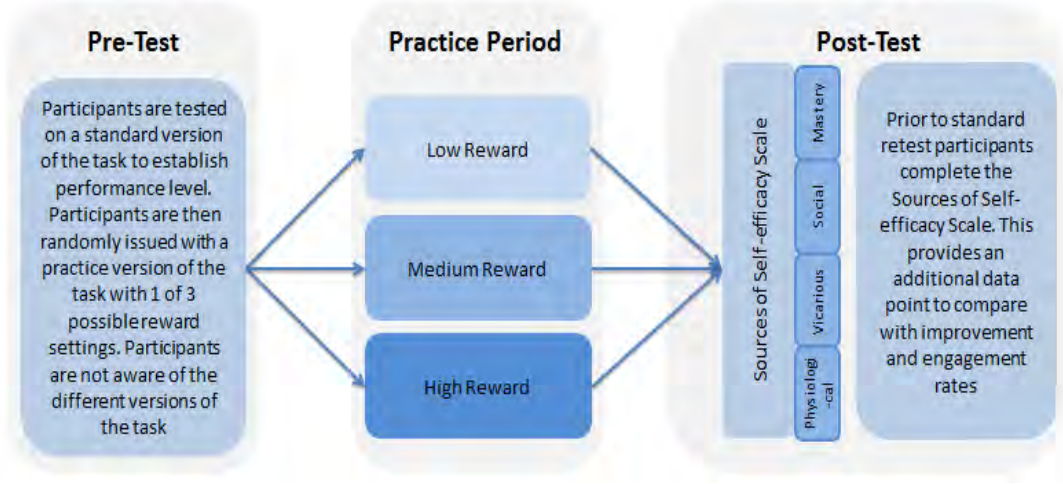

Figure 6. Round 5 Test Structure

Data Analysis

Each round of testing followed a consistent analysis strategy. This strategy can be broadly categorised as ascertaining comparability, examining the impact of the practice period and exploring the impact of the variable in question.

Within Round Comparisons

In order to ascertain within round comparability the pre-test scores of all participants were grouped, based on the version of practice task they received, and compared using an analysis of variance test (ANOVA). Given the relatively small sample size normality is examined numerically and graphically. The F-statistic based ANOVA test was selected due to the conservative impact of a violation of the normality or homogeneity of variance assumptions (Donaldson 1966). Tiku (1971) echoes this sentiment and suggests that results from F-statistic based tests on non-normal data differ result in conservative estimations of variance. This difference decreases sharply as sample size increases.

In order to compare the means of pre-test and post-test scores of each group a paired samples T-test (repeated measures) was utilised. In order to examine effect size the eta squared statistic was calculated for each group. In an effort to further examine the effect size of each group, and to enhance the comparability of the results through any future meta-analysis, Cohen's d was calculated using the pooled standard deviation as suggested by Rosnow and Rosenthal (1996). In order to allow for the growing popularity of Glasses' delta (Thompson, 2002), and possible future use in meta-analysis, the initial sample standard deviations have been included (Glass et al., 1981). A Pearson's analysis was employed in order to examine the relationship between engagement and performance.

Between Round Comparisons

Prior to comparisons between separate rounds of testing an Independent Samples T-test was used to investigate potential variance in pre-test scores. A non-significant result supports between round comparability. A significant result could still allow for meaningful comparisons to be made if the pre-test variations between rounds is controlled for. However for the purposes of this study this was not necessary. A wider ANOVA was carried out in order to examine the pre-test scores of all rounds post final data collection. This was carried out with the same intent as the previous T-test but in the wider context of a complete data set. Similar to the within round analysis strategy outlined previously a Pearson's correlation analysis was employed to examine the relationship between performance and engagement with the practice task.

Sources of Self-Efficacy Scale Data

The 'Sources of Self-Efficacy Scale' was originally designed for use in the domain of mathematics (Usher and Pajares 2009). The scale was employed in round 5 of this study as participants completed the pre and post-test using the computerised maze navigation task. In order for the scale to be used in this study minor alterations were made regarding the wording of certain items. Theoretically, the alteration of the domain of operation in this instance is supported. As noted by several leading researchers in the area the explanatory and predictive power of self-efficacy increases in line with the specificity of the domain of operation (Bandura 1997, Pajares 1997, Usher and Pajares 2008). Noting the more defined nature of the task in question, and the comparatively low level of prerequisite knowledge or skills, the new domain of operation is viewed as more constrained. Although the changes were relatively minor in nature Arndt and De Bruin (2011) note that even subtle

alterations can damage claims of validity from previously developed scales. Heeding this warning the validity of the adapted scale is not assumed. Acknowledging the small sample size in this application of the scale factor analysis methods would not be suitable for assessing validity. In order to examine the potential impacts of alterations all items will be analysed in terms of correlations. The relationships between items will then be examined in terms of alignment with their respective subscales and with all other items. This allows for a rudimentary examination of 'best fit' and will provide information regarding potential loading of items on multiple factors.

When examining data gathered using the scale source variances will be considered using difficulty groupings as nominal variables. This allows for the comparisons of mean source scores in the context of varying difficulty levels. A one-way between groups ANOVA was used to examine whether difficulty groups reported significantly different self-efficacy source levels. In order to examine the impact of difficulty on each reported source effect sizes were calculated from eta squared values. Post-hoc comparisons using the Tukey HSD test allow for further examinations of the degree to which source values varied relative to difficulty groups.

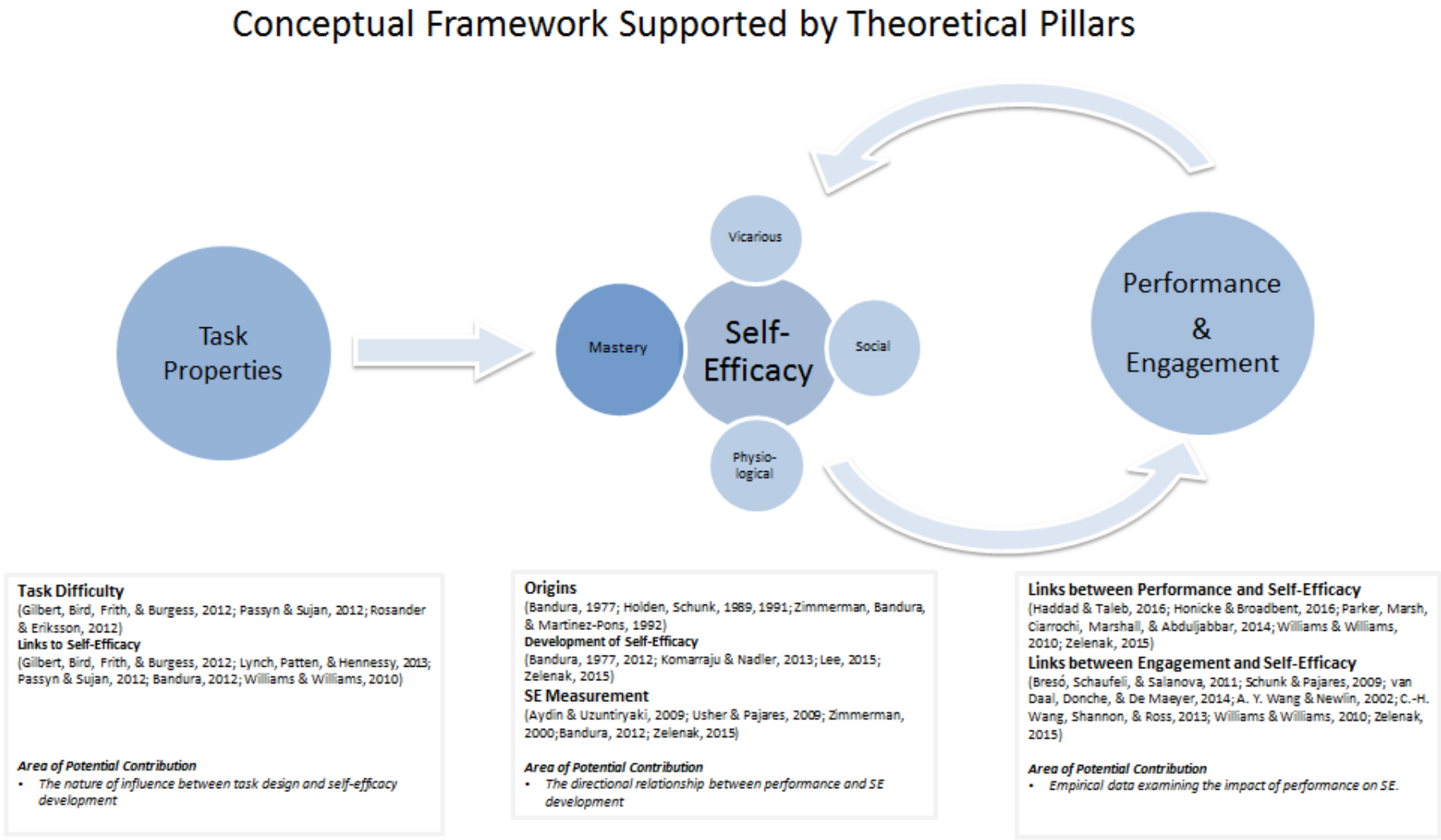

Figure 7. Conceptual Framework

Ethical Considerations

Research involving children must meet the highest ethical standards due to the vulnerability of the participants. Consequently, stringent ethical parameters have been constructed in order to protect participants from any negative impact. Ethical considerations have been negotiated with current participants. Participants have the right to withdraw from the study at any stage without prejudice. The four basic ethical principles of respect for persons, beneficence, non-malfeasance and justice (LoBiondo-Wood and Haber 1998) were prioritised at all times. In order to construct a robust ethical policy, standards have been set out according to the following key areas in alignment with the Irish Department of Children and Youth Affair ethical policy document (Department of Children and Youth Affairs, 2012). This project has also been granted ethical approval by the Faculty of Science & Engineering Research Ethics Committee, University of Limerick.

Minimising risk of harm

The perceived physical, social and psychological risk of participation in this study is viewed by the researcher as being relatively low. Although risk has been deemed low, a keen understanding exists for the underlying social risks that may vary in severity for each individual. Unforeseen risk may arise and accordingly a 'minimal risk' standard has been established, as supported by international research guidelines (Kopelman 2004). This standard equates to a level of risk to the participant that is less than the participant would experience in routine physical activity, daily life or psychological testing. Measures to mitigate any potential harm have been put in place. These include:

a) An independent contact in the form of the Chair of the Faculty Ethics Committee of the host institution.
b) A recommended screen time limit has been issued in accordance with international standards (Sigman 2007).
c) A silent "opt out" facility has been designed so that a child can communicate their desire not to participate privately reducing social risk. This would allow withdrawal to be non-evident to peers.

In the event that any of these standards are not met at any stage, and as a result level of potential risk exceeds that of 'minimal risk', the project will be suspended.

Informed consent and assent

In order to consider consent valid, it must also be informed (Shaw *et al.* 2011). It is the responsibility of the researcher to put in place reasonable mechanisms to ensure that the person whose consent is being sought has been provided with suitable information relating to the project. The onus is on the researcher to establish suitable means of communication to ensure that all parent/guardians and participants can raise any concerns with the researcher or alternatively an independent representative detailed above. Contact information will be supplied in accordance with the Data Protection Acts (Data Protection Commissioners, 2007). Guardians will be supplied with a comprehensive information pack including multiple contact options. This pack will also contain a consent form. Informed consent from the child will also be sought after information is presented in a child-appropriate manner. Opportunities to assimilate information, consult with others and ask questions will be presented

before the child is asked to decide whether or not to assent as per international guidelines (Thomas and O'kane 2006).

Confidentiality and anonymity

In accordance with the Data Protection Acts 1988 and 2003 (Government of Ireland, 1988 and 2003) no identifiable data will be disclosed to others without the express permission of the participants and respective parent/guardian. The project has been designed in a manner that limits the amount of initial identifiable data. Coding will be applied at the initial contact point which will allow for the separation of personal information from statistical data thus removing direct identifiers. Storing, collection and accessing of data will all be carried out exclusively by the researcher in order to minimise risk. Anonymity will be maintained in all publications ensuing from this research. The right of the participant to access relating data is acknowledged and participants will be informed as such.

Legal requirements and policy commitments

United Nations Convention on the Rights of the Child (UN, 1989) outlines the many rights of the child. This is viewed as the most critical set of standards and as such guides every element of this project. Article 12 and 13 outline a child's right to express his or her views. The right to access appropriate information (Article 17) is noted and has influenced the design of information and consent documents. The rights of parents and children, under the Constitution of Ireland (1937) and the relevant Data Protection Acts (Government of Ireland, 2003), involved in the study have also been noted and have informed the development of the methodology and project as a whole.

Children's participation in the research process

Children's participation in the study must be from an informed and assenting position. Methodological formats have been designed in a manner that seeks to balance the child's right to participate in the study with their right to protection. Where specific needs of children arise every effort will be made by the researcher to meet these needs. Before engaging with pupils the researcher will develop a suitable knowledge of the pupil population to ensure that any specific needs that require prior preparation can be met. On completion of the research project the findings will be presented to participants in a suitable format highlighting their contribution and resultant benefits. The researcher appreciates the dynamic relationship between a professional adult and child. As explored by Morrow and Richards (1996) the desire of the researcher to achieve a high level of participation may influence a child's decision.

CHAPTER FOUR: RESULTS

> "[F]rom a description of some phenomena to a description of something which produces it or is a condition for it."
>
> (Bhaskar 2010, p.11)

Introduction

The following chapter outlines the results of all five rounds of testing. All rounds are inspected in order to investigate their comparability. Each round is then examined individually. Initially an ANOVA is used to compare all three groups within the round using their pre-test scores in order to examine whether any difficulty group is statistically unique. Each group's pre and post-test scores are then examined using paired T-tests and an effect size is calculated using eta squared. ANOVA tests are also used in comparisons of post-test scores. In addition to the performance data round 5 has accompanying sources of self-efficacy scale data (Usher and Pajares 2009). Scale data is compared across difficulty groups in round 5 in order to examine whether previously posited influences of task difficulty on self-efficacy are reflected in the sources of self-efficacy scale data. The end of the results chapter focuses on comparisons of rounds in order to examine the impact of manipulated variables. Finally, limitations of the study are discussed in light of the previously outlined results. Prior to an examination of individual rounds an ANOVA test was used in order to establish the comparability of each round based on pre-test scores. This one way between group analysis of variance was utilised in order to assess the homogeneity and comparability of each round using pre-test scores. There was no statistically significant difference (p = 0.617) in the pre-test scores between the 5 rounds: F (4,312) = .665. This is supported by the relatively consistent pre-test

means (see Table 2). These results suggest homogeneity between rounds and supports comparability.

Table 2. Comparison of Pre-Test Means Across All Rounds

	N	Mean	σ
Round 1 Difficulty Variance Out of School	62	1824.84	1013.61
Round 2 Difficulty Variance In School	62	1741.77	912.9
Round 3 Proportional Reward	61	1886.72	878.58
Round 4 Reward Variance Only	66	1995	988.8
Round 5 Difficulty Variance with Scale	66	1906.48	822.95

When examining rounds, and groups within rounds, a labelling system was used with the round number (1, 2, 3, 4 or 5) followed by a letter (A, B or C). The letter corresponds to a level of the manipulated variable. For example group 2A is the group from round 2 that was issued with the low difficulty (the manipulated variable within round 2) version of the practice task. Conversely group 2C from round two practiced with the high difficulty version of the task. See Table 3 for a full list of all group labels.

Table 3. Group Labels

Round 1	**Difficulty Variable Out of School**
1A	Low Difficulty
1B	Medium Difficulty
1C	High Difficulty
Round 2	**Difficulty Variable In School**
2A	Low Difficulty
2B	Medium Difficulty
2C	High Difficulty
Round 3	**Proportional Reward**
3A	Low Difficulty/Low Reward
3B	Medium Difficulty/Medium Reward
3C	High Difficulty/High Reward
Round 4	**Reward Variable (Medium difficulty for all participants)**
4A	Low Reward
4B	Medium Reward
4C	High Reward
Round 5	**Difficulty Variable with Sources of Self-efficacy Scale**
5A	Low Difficulty
5B	Medium Difficulty
5C	High Difficulty

Results for Round 1 (Difficulty Variable Out-of-School)

A one-way between-group ANOVA was carried out to explore variances in the initial task performance between difficulty groups (1A, 1B and 1C) in round 1. This one way between group analysis of variance was utilised in order to assess the comparability of the population using pre-test scores. There was no statistically significant difference ($p = 0.716$) in the pre-test scores between the three difficulty

groups: F (2,59) = 0.336. This is supported by the very small effect size, calculated using eta squared, of 0.01 and the relatively consistent pre-test means (see Table 2). These results suggest homogeneity between groups 1A, 1B and 1C and supports comparability.

Table 4. Comparison of Means Round 1

	Mean	σ	Mean Difference	Improvement	t	Sig. (2-tailed)	η^2
Group1A							
Pre-test	1801.5	900.57	1748.5	97%	3.12	0.006	0.35
Post-test	3550	2874.66					
Group1B							
Pre-test	1958.2	1289	951.82	49%	3.016	0.007	.31
Post-test	2910	1761.85					
Group1C							
Pre-test	1701.5	860	451.5	27%	2.382	0.028	.24
Post-test	2153	671					

A subsequent mixed between-within (repeated measures) ANOVA was employed in order to assess the impact of the practice difficulty (High/Medium/Low) on participants improvement between pre and post-tests. The results suggest a non-significant difference in improvement between groups, Wilks-Lambda = .913, F (2,59) = 2.825, p = .067, eta squared = .087.

For round 1 the results show that on average group 1A, who practiced with the low difficulty task, demonstrated the greatest overall improvement between the pre-test and post-test. Group 1A demonstrated a mean improvement of 1749 points, equal to an improvement of 97% over the week. This resulted in a large effect size (Cohen 1988, p. 22), calculated using eta squared (η^2= 0.35).

Group 1B, who practiced with the medium difficulty task, demonstrated the second highest increase in mean scores with an increase of 951.9 points. This represented an average improvement of 49% in performance from the pre to post-test scores. This resulted in a large effect size calculated using an eta squared value (η^2= 0.31).

Group 1C, who practiced with the high difficulty version of the task, demonstrated the smallest overall improvement with a mean improvement of 451 points which represents a 27% improvement on the pre-test mean score. The magnitude of difference between the pre-test mean score and post-test mean score was found to be large (η^2= 0.24).

A one-way between-group ANOVA test which examined the amount of games played by each difficulty group during the practice period proved to be statistically significant F (2,59) = 4.138, p = 0.021. A subsequent Post-hoc comparison using the Tukey HSD indicated that group 1A (Low Difficulty: M=18.05, SD=10.175) was statistically significantly different to group 1C (High Difficulty: M=10.5, SD=8.056). Group 1B (Medium Difficulty: M=13.05, SD=6.959) did not differ significantly from either group 1A or 1C. Noting the considerable difference in means, and statistically different nature of groups 1A and 1C, the effect size was calculated using eta squared. A medium effect size was noted (η^2=0.12).

A subsequent Pearson's correlation analysis examining games played relative to improvement was found to be statistically significant (r=0.652, n=62, p<0.001), resulting in a large correlation (Cohen 1988). This suggests that level of engagement is an influential factor of improvement and also supports the findings of Lynch, Patten and Hennessy (2013).

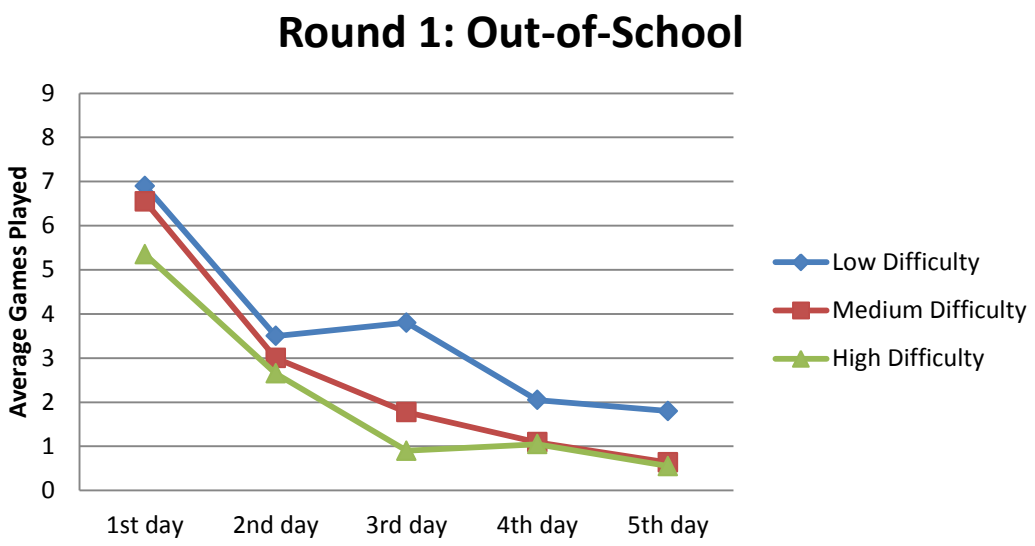

Figure 8. Engagement Rate of Round 1

The engagement rates over time per difficulty group are shown in Figure 8. A distinct higher engagement rate is noticeable for group 1A who were practicing with the low difficulty task. A similar reduction in engagement towards the end of the practice period was witnessed for groups 1B and 1C who practiced using the medium and high difficulty tasks respectively. These groups demonstrated a marked drop in engagement on the final day. The difference in engagement rates observed in Figure 8 should be considered in conjunction with the previously outlined correlation between engagement rate and improvement (r=0.652, n=62, p<0.001).

The results of round 1 demonstrate a considerable difference in engagement and performance scores across difficulty groups. This highlights a relationship between task difficulty and performance. Those practicing with the low difficulty version of the practice task (group 1A) demonstrated a consistently higher rate of engagement across the practice period and an associated increased improvement rate over groups 1B and 1C. This pattern continues with those practicing with the medium difficulty (group 1B) task demonstrating the second highest engagement rate and overall performance. Group 1C demonstrated both the lowest engagement rate and overall

performance. Difficulty variations corresponded to observed variations in engagement and performance.

Results for Round 2 (Difficulty Variable In-School)

The purpose of this round was to examine whether the impact of task difficulty on engagement and performance observed in the previous round was consistent in a school setting. This round used an identical testing procedure and task versions as the previous round. Participants were pre-tested using a standard version of the task to ascertain their initial performance levels. They were then issued with one of three practice versions of the task. The three versions of the task were defined by their difficulty settings. Participants had no knowledge of these difficulty variations. This resulted in three groups within the round. Group 2A who were issued with the low difficulty version of the task, group 2B who were issued with the medium difficulty version of the task and group 2C who were issued with the high difficulty version of the task. Participants used their assigned version of the task of their own volition throughout the practice period unaware that their peers could be using an alternative version to their own. Once the five day practice period was complete participants were retested using the same standard version of the task used in the pre-test. Rate of improvement was calculated by comparing pre and post-test scores. Engagement rate data was also recorded on the USB device issued to participants at the start of the practice period and collected after the post-test.

A one-way between-group ANOVA was carried out to explore variances in the initial task performance between difficulty groups in the Out-of-School round. There was no statistically significant difference ($p = 0.869$) in the pre-test scores between difficulty groups: F (2,59 = 0.141). This is supported by the small effect size,

calculated using eta squared of 0.005, and the relatively small differences between means of pre-test scores (see Table 5). This suggests homogeneity across all difficulty groups and supports comparability.

Table 5. Comparison of Means Round 2

	Mean	σ	Mean Difference	Improve-ment	t	Sig. (2-tailed)	η^2
Group 2A							
Pre-test	1689.05	1001.57	1013.1	60%	3.524	0.002	0.384
Post-test	2702.15	1789.92					
Group 2B							
Pre-test	1828.57	959.04	346.57	19%	1.605	0.124	0.114
Post-test	2175.14	534.53					
Group 2C							
Pre-test	1706	799.23	273	16%	1.537	0.141	0.11
Post-test	1979	567.73					

A subsequent mixed between-within (repeated measures) ANOVA was employed in order to assess the impact of the practice difficulty (High/Medium/Low) on participants improvement between pre and post-tests. The results suggest a non-significant difference in improvement between groups, Wilks-Lambda = .906, F (2,59) = 3.074, p = .054, eta squared = .094.

Results show that group 2A, who practiced with the low difficulty version of the task, demonstrated the greatest overall improvement between the pre-test and post-test. Group 2A demonstrated a mean improvement of 1013.15 points or 60%. This

resulted in a large effect size, calculated using eta squared, of 0.384 (Cohen 1988, p.22).

Group 2B, who practiced with the medium difficulty version of the task, demonstrated the second highest increase in mean scores with an increase of 905. This represented a 19% increase in the mean post-test score. This again resulted in a large effect size calculated using an eta squared value of 0.27.

Group 2C, who practiced with the high difficulty version of the task, demonstrated the smallest overall improvement with a mean improvement of 361 which represents a 22% improvement on the pre-test mean scores. The eta squared statistic of 0.157 is still large (Cohen 1988, p.22) but is notably smaller than the effect size reported for both groups 2A and 2B.

A final one-way between-group ANOVA was conducted in order to examine engagement rates (See Figure 9.) across groups as a possible cause of the variance in post-test scores between groups. The results indicate statistically significant difference (p = 0.013) in engagement rates across groups. This effect is attributed to the hidden difficulty variations of the practice tasks. A subsequent Post-hoc comparison using the Tukey HSD indicated that group 2A (Low Difficulty: M=15.33, SD=10.08) was statistically significantly different to group 2C (High Difficulty: M=6.95, SD=5.643). Group 2B (Medium Difficulty: M=11.62, SD=9.86) did not differ significantly from either group 2A or 2C. Noting the considerable difference in means, and statistically significant difference between groups 2A and 2C, the effect size was calculated using eta squared. A medium effect size was noted (η^2=0.14) (Cohen 1988).

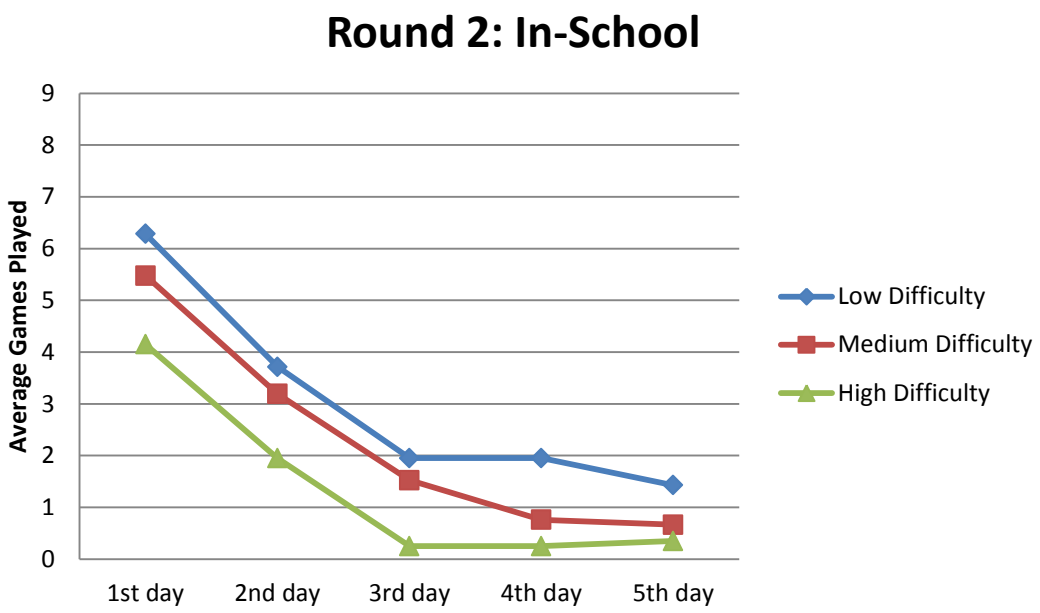

Figure 9. Engagement Rate of Round 2

The engagement rate of each group (2A, 2B and 2C) are graphically represented in Figure 9. The results demonstrate a higher practice rate for group 2A participants who were issued with the low difficulty task. Group 2C who practiced with the high difficulty version of the task demonstrate an early disengagement with average per day attempts falling below one by the third day of the practice period. The engagement rates illustrated above reflect the performance variation previously discussed. This should be considered in the context of a Pearson's correlation test which examined engagement rate relative to improvement. A large and statistically significant correlation was observed (r=0.602, n=62, p<0.001) (Cohen 1988).

The results from round 2 present a clear pattern when compared across groups. This pattern is similar in nature to round 1 with difficulty levels being inversely proportional to engagement and performance. Group 2C who practiced with the high difficulty task demonstrated the lowest engagement and performance figures. Conversely group 2A who practiced with the low difficulty task again demonstrated the highest engagement rates and performance increases. Although the nature of this

pattern is similar to round 1 data, key differences should be noted. The only difference in task conditions between the two rounds was the practice environment. While the pattern between difficulty groups remained consistent the overall performance between groups demonstrated considerable mean differences. Round 1 post-test scores (see Table 4) were noticeably higher than those observed in round 2 (see Table 5).

Comparing the Results of Round 1 and Round 2

A previous analysis of variance test examining pre-test scores of all 5 rounds showed no statistically significant difference in pre-test scores between rounds (see Table 2.). This suggests that the two rounds compared in this section had equivalent initial ability on this computerised maze navigation task.

Round 1 (M= 2872.42, SD = 2036) outperformed round 2 (M = 2290.37, SD = 1157.25) during the post-test. In an effort to explore the cause of the differences in performance witnessed between these two rounds, and difficulty groups within these rounds, further testing and comparisons were employed. Difference in post-test scores between round 1 and 2, and between the respective sub-groups, can be seen in Figure 10.

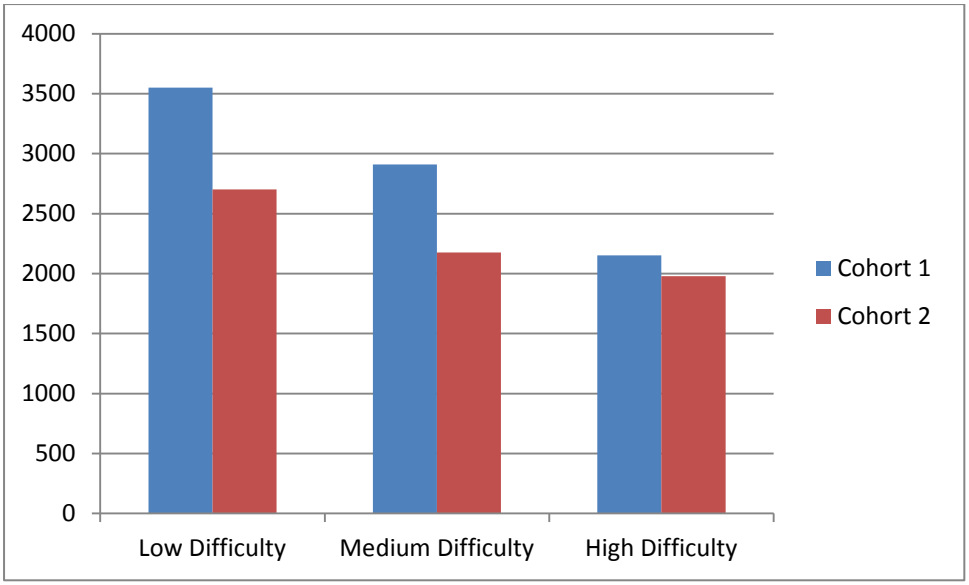

Figure 10. Comparison of Difficulty Groups Post-test Scores across Round 1 and Round 2

Participants who practiced with the lowest difficulty version showed the greatest overall improvement and conversely those who practiced using the highest difficulty version showing the lowest overall improvement. While this trend is evident in both rounds, it is clearly more prominent within round 1 (Out-of-School).

When post-test scores from each round are compared the inverse relationship between the difficulty level of the task and participants' level of improvement is evident. Groups 1A and 2A, who practiced with the easiest version of the game, scoring highest within both rounds. The magnitude of difference between the mean scores for participants practicing with the easier version of the task was considerably larger compared to those practicing with the standard and harder version across both rounds (see Table 4 and Table 5). Participants who practiced with the standard version of the task also demonstrated large effect sizes across both round 1 (η^2 = 0.107) and round 2 (η^2 = 0.27). Finally, participants who practiced with the harder version of the task demonstrated a moderate effect size for round 1 (η^2 = 0.099) and a large effect size for round 2 (η^2 = 0.157).

The higher levels of improvement demonstrated by round 1 are related to participants' increased levels of engagement with the task over the practice period when compared to round 2. As can be seen in Figure 11 on average round 1 participants completed more practice tasks over the week than participants from round 2. A Pearson's correlation analysis suggests that a statistically significant large correlation exists between number of tasks completed in practice (level of engagement) and levels of improvement for Round 1 (r=0.652, n=62, p<0.001) (Cohen 1988). A similarly large and statistically significant relationship was observed between engagement rate and improvement for round 2 (r=0.602, n=62, p<0.001). Variance in group improvement rates and engagements rates are potentially due to the altered environments. This will be explored further in the discussion section.

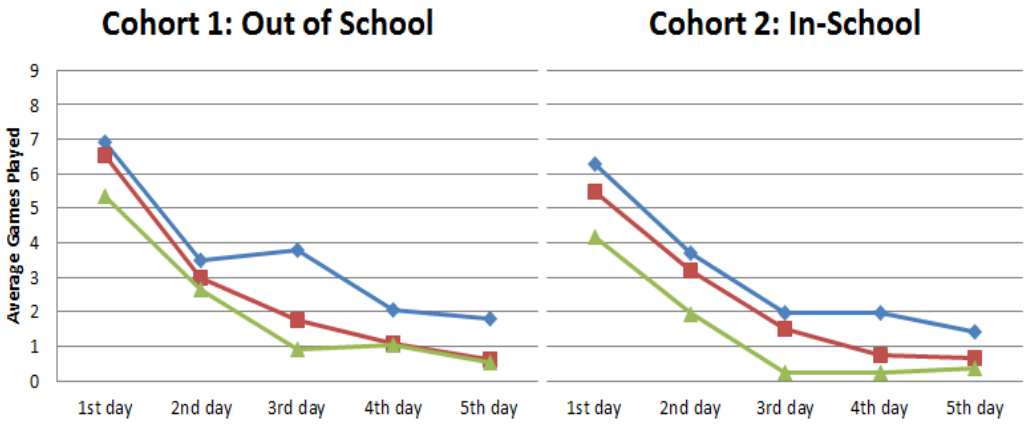

Figure 11. Comparison of Mean Games Played Between Rounds 1 and 2

Results for Round 3 (Proportional Reward)

The previous two rounds suggested task difficulty exerted a similar influence on engagement and performance in disparate settings. In order to gauge this effect, a proportional reward system was used. In this manner the degree to which a manipulated reward setting counteracted a manipulated difficulty setting has the potential to provide an indication of how an individual assessed their own

performance. Those practicing with the high difficulty version of the task, previously linked to lower engagement and performance levels, were provided with a high reward system. This rewarded these participants with more 'points' for each positive action within the task. Conversely, those practicing with the low difficulty level task, previously linked to increased engagement and performance, were provided with a lower reward system which awarded participants with fewer 'points' for each positive action. Those practicing with the medium difficulty retained the standard reward system. On all versions of the task in all rounds 'points' are displayed in a score window in the bottom part of the display.

A one-way between-group ANOVA was carried out to explore variances in the initial task performance between groups (3A, 3B and 3C) in Round 3. This one way between group analysis of variance was utilised in order to assess the homogeneity and comparability of the population using pre-test scores. There was no statistically significant difference (p = 0.686) in the pre-test scores between the three difficulty groups: F (2,58) = 0.38. This is supported by the very small effect size, calculated using eta squared, of 0.013 and the relatively consistent pre-test means (see Table 6). These results suggest homogeneity between groups 3A, 3B and 3C and supports comparability.

Table 6. Comparison of Means Round 3

	Mean	σ	Mean Difference	Improvement	t	Sig. (2-tailed)	η^2
Group 3A							
Pre-test	1997	744.843	3279	164%	3.066	.006	0.33
Post-test	5276	4941.529					
Group 3B							
Pre-test	1755	732.526	1913.5	109%	1.718	.057	0.177

Post-test	3668.5	4261.769					
Group 3C							
Pre-test	1907.14	1117.552	1340.48	70%	2.69	.014	0.266
Post-test	3247.62	2657.51					

A subsequent mixed between-within (repeated measures) ANOVA was employed in order to assess the impact of the proportional reward system (High/Medium/Low) on participants improvement between pre and post-tests. The results suggest non-significant difference in improvement between groups, Wilks-Lambda = .955, F (2,58) = 1.382, p = .259, eta squared = .045.

For round 3 the results show that on average group 3A, who practiced with the low difficulty/low reward version of the task, demonstrated the greatest overall improvement between the pre-test and post-test. Group 3A demonstrated a mean improvement of 3279 points, equal to an improvement of 164% over the week. This resulted in a large effect size (η^2= 0.33) (Cohen 1988, p.22).

Group 3B, who practiced with the medium difficulty/medium reward version of the task, demonstrated the second highest increase in mean scores with an increase of 1913.5 points. This represented an average improvement of 109% in performance from their pre to post-test scores. This resulted in a large effect size calculated using an eta squared value (η^2= 0.192).

Group 3C, who practiced with the high difficulty/high reward version of the task, demonstrated the smallest overall improvement with a mean improvement of 1340.5 points which represents a 70% improvement on the pre-test mean score. The magnitude of difference between the pre-test mean score and post-test mean score was again found to be large (η^2= 0.34).

An ANOVA test which examined the amount of games played by each difficulty group proved not to be statistically significant, F (2,58) = 1.644, p = 0.202. This is in contrast to the previous 2 rounds and suggests the increased reward rate used for those practicing with the higher difficulty task, and conversely the lower reward rate used for those practicing with the lower difficulty task had a normalising affect on engagement (See Figure 12). A subsequent Pearson's correlation analysis examining games played relative to improvement proved also to be statistically significant (r=0.498, n=61, p<0.001), resulting in a large correlation (Cohen 1988). This supports the supposition that level of engagement is an influential factor of overall improvement.

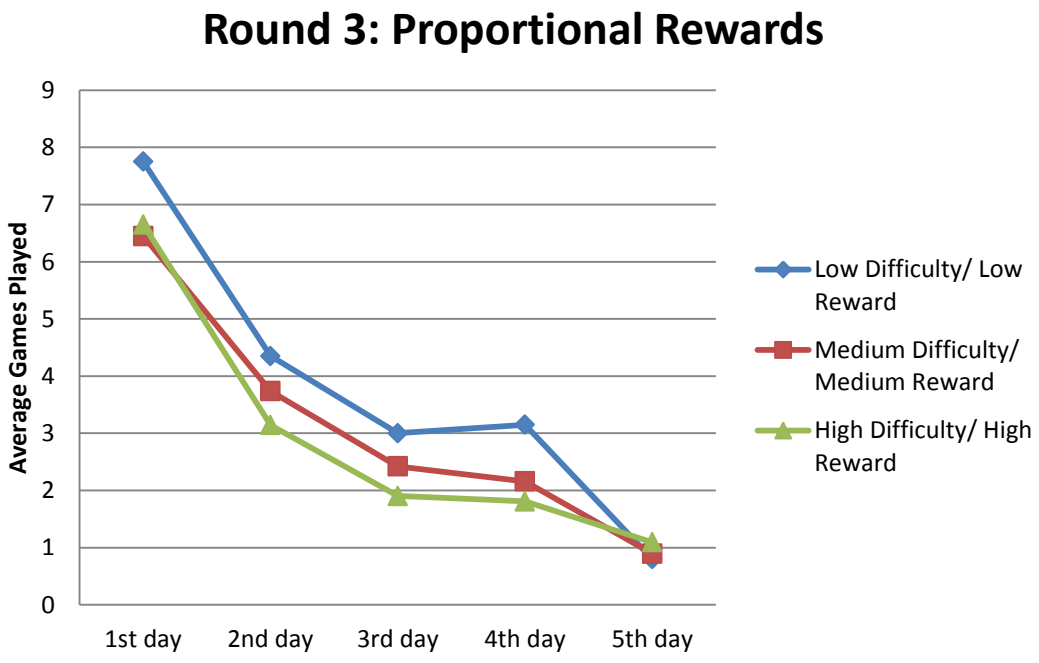

Figure 12. Engagement Rate of Round 3

The engagement rates of groups 3A, 3B and 3C are graphically represented in Figure 12. In line with the previous performance data group 3A, who were issued the low difficulty/ low reward version of the task, demonstrated higher engagement rates throughout the practice period. Group 3B, who were issued the medium

difficulty/medium reward version of the practice task, displayed the second highest engagement rate. Group 3B's engagement rate is closer to that of group 3C who were issued with the high difficulty/ high reward version of the task. Group 3C displayed the lowest overall engagement rate and the overall lowest performance improvements as noted previously.

Comparing the Results of Round 3 to Rounds 1&2

The data from round 3 maintains the inverse pattern of difficulty relative to engagement and performance observed in round 1 and 2. The normalising effect of the proportional reward system that this round of testing was designed to investigate was not observed. Conversely, the overall mean improvement scores were substantially higher than any preceding round. An ANOVA test demonstrated a statistically significant difference in post-test scores between rounds 1, 2 and 3, $F(2,182) = 6.751$, $p = 0.001$. A subsequent Tukey HSD posthoc test indicated that round 3 (M=4050.66, SD=2787.9) was statistically significantly different to round 1 (M=2872.42, SD=2036.07) and round 2 (M=2290.37, SD=1157.25). Round 1 and 2 demonstrated no statistically significant difference in post-test scores. In contrast to the predicted normalising effect on post-test scores, the proportional reward system led to considerably enhanced engagement and performance rates across all groups within round 3.

The predicted impact of manipulated reward is not evident when comparing between groups within the round. However, when round 3 is compared to previous rounds 1 and 2 the overall increase in performance and engagement is evident. This is supported by the previous ANOVA and is also apparent when means are compared (see Table 4, Table 5 & Table 6). The degree to which the post-test scores exceeded the previous two rounds, in addition to the lack of a negative impact of reward,

suggests a potential interaction effect between the reward and difficulty variables. Theoretically, this could be attributed to more favourable comparisons between all participants. This informed the development of the subsequent round.

Results for Round 4 (Reward Variable)

A one-way between-group ANOVA was carried out to explore variances in the initial task performance between groups 4A, 4B and 4C in round 4. This one way between group analysis of variance was utilised in order to assess the homogeneity and comparability of the population using pre-test scores. There was no statistically significant difference (p = 0.791) in the pre-test scores between the three difficulty groups: F (2,63) = 0.235. This is supported by the very small effect size, calculated using eta squared, of 0.007 and the relatively consistent pre-test means (see Table 7). These results suggest homogeneity between participant groups 4A, 4B and 4C and supports comparability.

Table 7. Comparison of Means Round 4

	Mean	σ	Mean Difference	Improvement	t	Sig. (2-tailed)	η^2
Group 4A							
Pre-test	1892.27	1122.972	435.91	23%	2.918	0.011	0.268
Post-test	2328.18	1250.922					
Group 4B							
Pre-test	1993.64	698.785	900.45	45%	2.909	0.008	0.287
Post-test	2894.09	1786.192					
Group 4C							
Pre-test	2099.09	1120.046	1220	58%	3.464	0.002	0.364
Post-test	3319.09	2234.594					

A subsequent mixed between-within (repeated measures) ANOVA was employed in order to assess the impact of reward level (High/Medium/Low) on participants

improvement between pre and post-tests. The results suggest a non-significant difference in improvement between groups, Wilks-Lambda = .943, F (2,63) = 1.907, p = .157, eta squared = .057.

For round 4 the results show that on average group 4C, who practiced with the high reward version of the task, demonstrated the largest overall improvement between the pre-test and post-test. Group 4C demonstrated a mean improvement of 1220 points, equal to an improvement of 58% over the week. This resulted in a large effect size (Cohen 1988, p.22), calculated using eta squared, (η^2= 0.36).

Group 4B, who practiced with the medium reward version of the task, demonstrated the second highest increase in mean scores with an increase of 900.5 points. This represented an average improvement of 45% in performance from pre to post-test scores. This resulted in large effect size calculated using an eta squared value (η^2= 0.287).

Group 4A, who practiced with the low reward version of the task, demonstrated the smallest overall improvement with a mean improvement of 436 points which represents a 23% improvement on the pre-test mean score. The magnitude of difference between the pre-test mean score and post-test mean score was found to be large (η^2= 0.268).

An ANOVA test which examined the amount of games played by each group proved not to be statistically significant, F (2,63) = 1.843, p = 0.167. A Pearson's correlation analysis examining games played relative to improvement proved to be statistically significant (r=0.494, n=66, p<0.001), resulting in a large correlation (Cohen 1988). This supports the supposition that level of engagement is an

influential factor of overall improvement and again aligns with the findings of Lynch, Patten and Hennessy (2013).

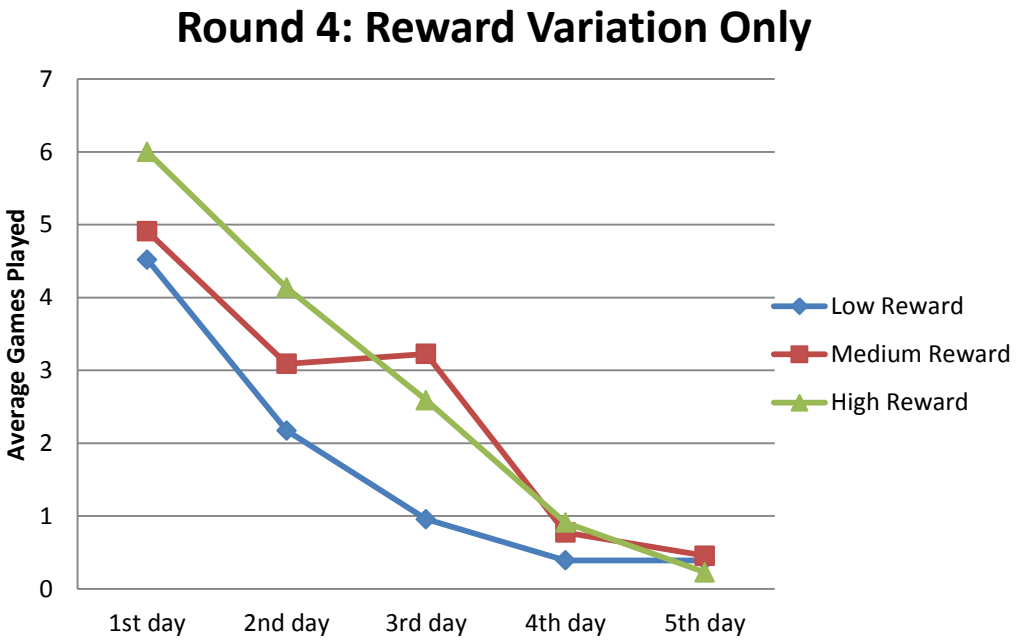

Figure 13. Engagement Rate of Round 4

The engagement rate of groups 4A, 4B and 4C are graphically represented in Figure 13. In contrast to previous rounds the distinction between groups, particularly between group 4C (high reward) and group 4B (medium reward) demonstrates considerable overlap towards the later stage of the practice period. This is reflected in the comparatively small variations in performance figures previously outlined. Round 4 data demonstrated considerably lower post–test scores than round 3 supporting suspicions of an interaction effect between reward and difficulty in the previous round. Group 4C who practiced with the low reward version of the task demonstrated the lowest performance increase and lowest engagement rate. Conversely group 4A, who practiced with the high reward version of the task, demonstrated the highest performance and engagement rates.

The influence of reward, when examined between groups, was considerably lower than the influence of difficulty observed in rounds 1, 2 or 3. The amount of games played by each group between the round proved not to be significantly different F (2,63) = 1.843, p = 0.167. In addition, the differences in effect size and post-test means were less than those observed in previous rounds (see Table 7). This suggests that reward influence is less than that of difficulty. However, the results of round 3 raise questions relating to the interaction of difficulty and reward.

Comparing the Results of Round 3 and Round 4

A previous analysis of variance test examining pre-test scores of all 5 rounds showed no statistically significant difference in pre-test scores between rounds (see Table 2.). This suggests that the two rounds had an equivalent initial ability on this computerised maze navigation task. However, a statistically significant difference was witnessed in the post-test scores of both rounds. An independent-sample T-test demonstrated that round 3 (M= 4050.66, SD = 4075.92) significantly outperformed Round 4 (M = 2847.12, SD = 1821.23) during the post-test; t (125) = 2.176, p = 0.031.

In an effort to explore the cause of the differences in performance witnessed between these two rounds, and groups within these rounds, further testing and comparisons were employed. Difference in post-test scores between round 3 and 4, and between the respective sub-groups, can be seen in Figure 14 and Figure 15 and is highlighted by the relative uniformity of the pre-test scores for each round (see Table 2). The group within Round 3 that practiced with the low difficulty version of the task and proportional low reward level of the task scored highest. As the group's difficulty levels increased their performance levels decreased (See Figure 14.).

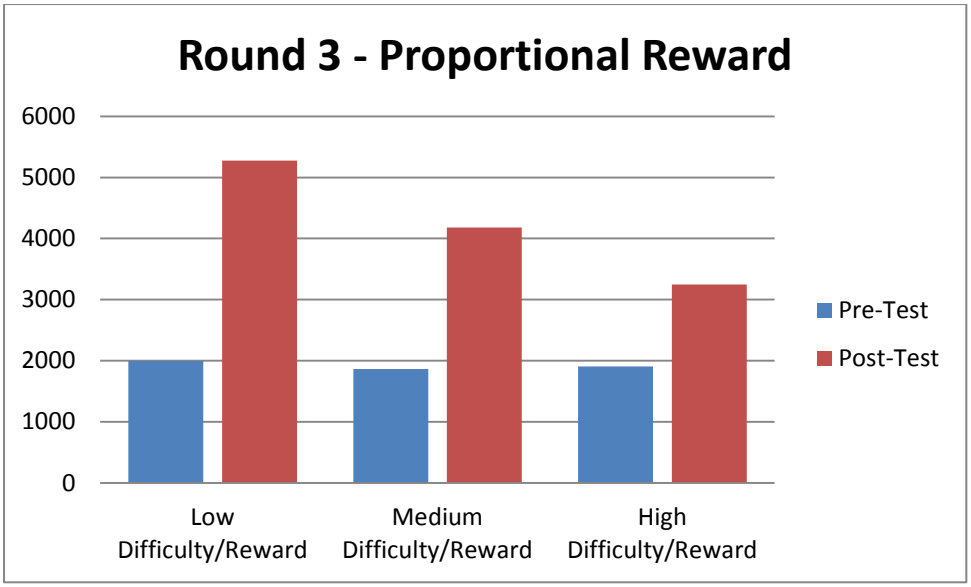

Figure 14. Comparison of Pre-Test and Post-Test Scores of Round 3

For Round 4 participants who practiced with the high reward version showed the greatest overall improvement and conversely those who practiced using the low reward version showed the lowest overall improvement (see Figure 15).

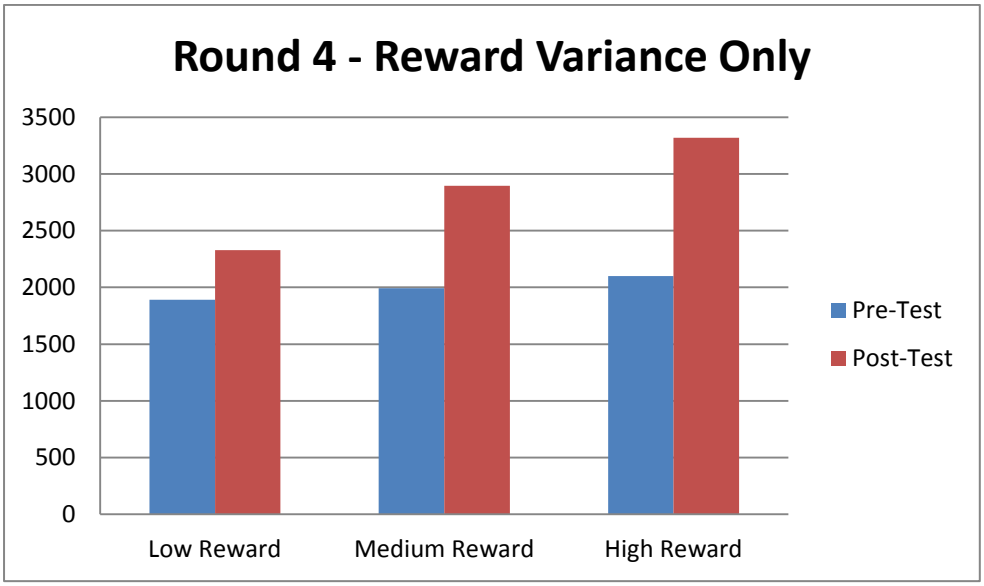

Figure 15. Comparison of Pre-Test and Post-Test Scores of Round 4

On average round 3 participants completed more practice (M = 16.49) tasks over the week than participants from round 4 (M = 11.77). A Pearson's correlation analysis

examining round 3 engagement and improvement rates (r=0.503, n=61, p<0.01) demonstrated a large statistically significant correlation. The same analysis using Round 4 data showed a statistically significant medium correlation (r=0.374, n=67, p<0.01). A graphical comparison of engagement rates is shown in Figure 16. The difference in overall engagement between groups is evident. This will be explored further in the discussion section.

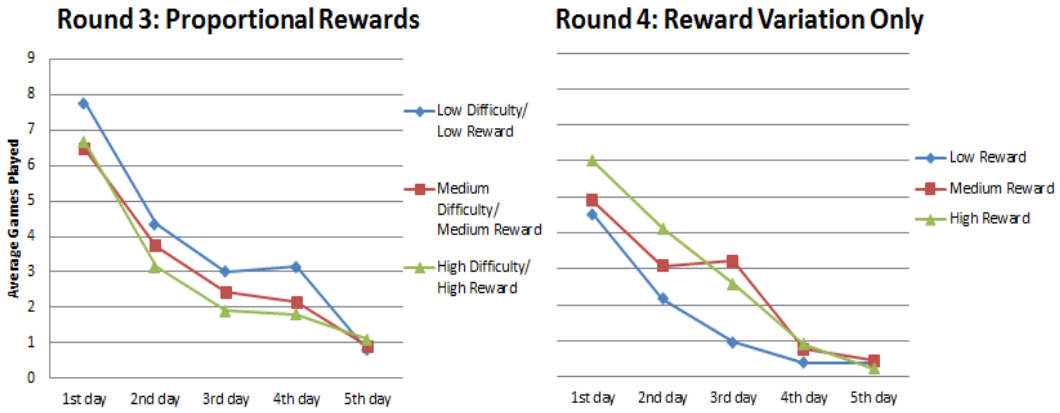

Figure 16. Comparison of Mean Games Played Between Rounds 3 and 4

Results for Round 5 (Difficulty Variable)

A one-way between-group ANOVA was carried out to explore variances in the initial task performance between difficulty groups (5A, 5B and 5C) in round 5. This one way between group analysis of variance was utilised in order to assess the homogeneity and comparability of the population using pre-test scores. No statistically significant difference (p = 0.425) in the pre-test scores between the three difficulty groups was identified: F (2,63) = 0.868. This is supported by the very small effect size, calculated using eta squared, of 0.027 and the relatively consistent pre-test means (see Table 8). These results suggest homogeneity of participant groups 5A, 5B and 5C and supports comparability.

Table 8. Comparison of Means Round 5

	Mean	σ	Mean Difference	Improvement	t	Sig. (2-tailed)	η^2
Group 5A							
Pre-test	1887.27	916.89	1465.91	78%	4.78	.001	0.52
Post-test	3353.18	1883.47					
Group 5B							
Pre-test	2079	850.69	855.09	41%	4.21	.001	0.46
Post-test	2934.09	888.26					
Group 5C							
Pre-test	1753.18	689.7	426.82	24%	2.69	.014	0.34
Post-test	2180	711.15					

A subsequent mixed between-within (repeated measures) ANOVA was employed in order to assess the impact of practice difficulty (High/Medium/Low) on participants improvement between pre and post-tests. The results suggest a statistically significant difference in improvement between groups, Wilks-Lambda = .864, F (2,63) = 4.942, p = .01, eta squared = .136.

For round 5, the results show that on average group 5A demonstrated the greatest overall improvement between the pre-test and post-test. Group 5A demonstrated a mean improvement of 1465.9 points, equal to an improvement of 78% over the week. This resulted in a large effect size (Cohen 1988, p.22), calculated using eta squared (η^2= 0.52).

Group 5B demonstrated the second highest increase in mean scores with an increase of 855.09 points. This represented an average improvement of 41% in participant performance from their pre to post-test scores. This resulted in a large effect size calculated using an eta squared value (η^2= 0.458).

Group 5C demonstrated the smallest overall improvement with a mean improvement of 426.81 points which represents a 24% improvement on the pre-test mean score. The magnitude of difference between the pre-test mean score and post-test mean score was again found to be large (η^2= 0.34).

An ANOVA of games played by each group found a significant difference, F (2,63) = 5.54, p = 0.006 (Cohen 1988, p.22). A subsequent Post-hoc comparison using the Tukey HSD indicated that group 5A (Low Difficulty: M=16.23, SD=10.1) was statistically significantly different to group 5C (High Difficulty: M=8.45, SD=4.925). Group 5B (Medium Difficulty: M=12.18, SD=12.18) did not differ significantly from either group 5A or 5C. Noting the considerable difference in means, and statistically different nature of groups 5A and 5C, the effect size was calculated using eta squared. A medium effect size was noted (η^2=0.15).

A subsequent Pearson's correlation analysis examining games played relative to improvement was also statistically significant (r=0.734, n=66, p<0.001), resulting in a large correlation (Cohen 1988, p.22). This supports the findings of the previous 4 rounds. An additional ANCOVA highlights that there is a non-significant difference in improvement (post-test minus pre-test) when engagement rates are controlled for F(2,63) = .337, p = .715. This suggests that difficulty variations influenced performance increases by acting on engagement. Associated variations relating to source values and their relationship to performance and engagement are further explored in the following sections.

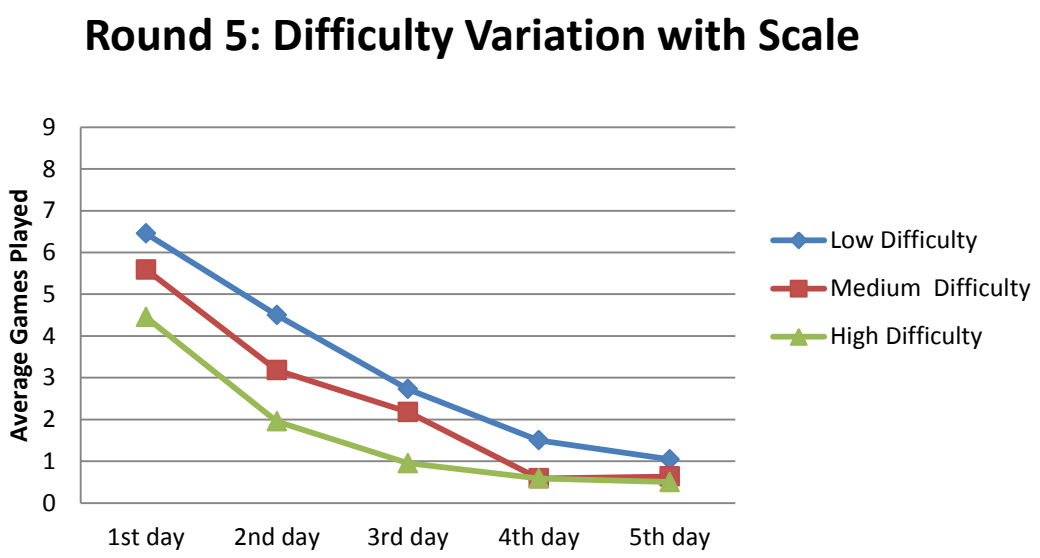

Figure 17. Engagement Rate of Round 5

The engagement rates of groups 5A, 5B and 5C are graphically represented in Figure 17 above. Group 5A, who practiced with the low difficulty version, demonstrated the greatest overall engagement rate. Group 5B, who were issued the medium difficulty practice task demonstrated the second highest engagement rate. Group 5B's engagement rate converged with that of 5C, who were issued the highest difficulty version of the task, toward the later stages of the practice period. The engagement rates shown mirror the performance data discussed previously.

Round 5 data echoes observations of difficulty variations on performance and engagement rates observed in previous rounds and data reported in (Lynch *et al.* 2013). Group 5A who practiced with the low difficulty version of the task demonstrated the greatest overall performance increase and highest engagement rates. Conversely group 5C demonstrated the lowest performance and engagement figures. The associated sources of self-efficacy scale results are discussed in a later section. The links between engagement rates and performance observed in previous rounds are consistent with the results of this round. The influence of task difficulty

on engagement is demonstrated across all difficulty groups. This also echoes the results of the previous rounds. An overview of all pre-test, post-test and resultant eta squared values are presented in Table 9. The eta squared values shown should be considered in terms of the standard deviation of each sample. Due to the relatively small sample size variance in the standard deviation of groups is expected but not desirable.

Table 9. Combined Pre-test, Post-test and Effect Sizes

			Pre-Test		Post-Test		
		N	Mean	σ	Mean	σ	η^2
Round 1 Difficulty Variance out of school	Low	20	1801.5	900.569	3550.5	2874.661	0.339
	Medium	22	1958.18	1259.666	2910	1753.022	0.31
	High	20	1701.5	834.59	2153	786.686	0.24
Round 2 Difficulty Variance in school	Low	21	1689.05	1001.568	2702.15	1789.922	0.384
	Medium	21	1828.57	959.037	2175.14	534.525	0.27
	High	20	1706	799.233	1979	567.728	0.157
Round 3 Proportional Reward and Difficulty	Low	20	1997	744.843	5276	4941.529	0.33
	Medium	20	1755	732.526	3668.5	4261.769	0.177
	High	21	1907.14	1117.552	3247.62	2657.51	0.266
Round 4 Reward Variance Only	Low	22	1892.27	1122.972	2328.18	1250.922	0.268
	Medium	22	1993.64	698.785	2894.09	1786.192	0.287
	High	22	2099.09	1120.046	3319.09	2234.594	0.364
Round 5 Difficulty Variance with Survey	Low	22	1887.27	916.89	3353.18	1883.473	0.52
	Medium	22	2079	850.687	2934.09	888.26	0.458
	High	22	1753.18	689.702	2180	711.149	0.34

Results of Sources of Self-efficacy Scale used in Round 5

The sources of self-efficacy scale (Usher and Pajares 2009) was used in order to examine whether task difficulty influenced self-efficacy formation for the given task. Previous behavioural outcomes (engagement and performance) attributed to varying difficulty are theoretically linked to varying levels of self-efficacy (Schunk and Pajares 2009). In order to identify if task difficulty influenced self-efficacy the four sources of self-efficacy (mastery, social persuasions, vicarious experience and physiological state) were examined. Before presenting the post-test results of each round, data relating to reliability is examined. In addition the pre-test results are explored. This provide a baseline indication of self-efficacy for the given task. Due to the limited sample size more advanced statistical methods such as factor analysis could not be employed (Osborne and Costello 2009).

Pre-Test Results of Scale

In order to assess the homogeneity and comparability of participants within this round, pre-test results of the sources of self-efficacy scale data are compared. This data is presented graphically in Figure 18. Each pre-test source value is compared between difficulty groups using between groups analysis of variance. The results of these ANOVAs (See Table 10.) for each source are considered independently, and in conjunction with the associated post-test results, in later sections.

Table 10. Pre-Test Sources of Self-Efficacy Scale Results

Source	*F*	Sig.
Mastery	.392	.677
Vicarious Experience	3.093	.52
Social Persuasions	.031	.97
Physiological State	.197	.822

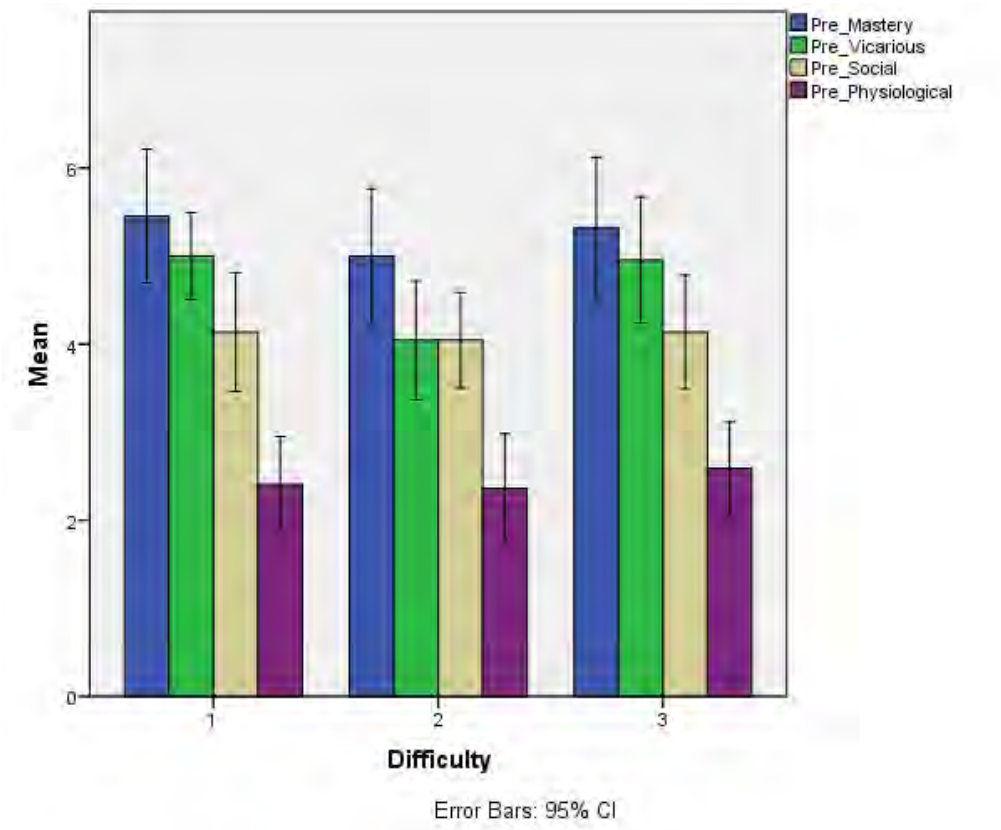

Figure 18. Pre-Test Sources of Self-Efficacy Scale Results

The results of the previously mentioned ANOVAs, in addition to the relatively consistent pre-test sources of self-efficacy means (See Figure 18.), suggest homogeneity and support comparability.

Post-Test Results of Scale

The internal consistency of all four subscales of the sources of self-efficacy scale (Usher and Pajares 2009), each containing six items, was examined and all were found to be above the recommended Cronbach's Alpha threshold of .80 (Henson 2001) (Mastery experience .973 vicarious experience .957 social persuasion 9.45 physiological state .929). As outlined by Bandura (1997) it is unusual for an individual to rely solely on one source when creating an estimate of their own self-efficacy in a given task. The inter-relatedness of the sources originally proposed by

Bandura is supported by the findings of this study (see Table 11). In the absence of a factor analysis, due to the small sample size, the correlation table shown can be used to interpret the relationship between items and their respective subscales. Noting the relatively high internal consistency, in conjunction with consistent individual item correlations, the resulting data can be examined under the reasonable assumption that the validity of the scale was not substantially damaged by the adaptation. The original scale and outlined minor alterations are included in Appendix B: Sources of Self-efficacy Scale. The darker highlighted sections of the matrix outline the degree of correlations between items that belong to the same source. The diagonal bold correlation figure which typically reads as 1, representing an items perfect correlation to itself, is instead replaced by the correlation of the individual item to the total source score. This is in line with the original validation study conducted by Usher and Pajares (2009).

The results of the scale are outlined in Figure 19 outline variance in each source attributed to varying difficulty levels. Predicted patterns were observed in Mastery Experience, Social Persuasions and less so in Physiological state. Physiological State was a negative report and as such displays an inverse pattern. Vicarious Experience demonstrated no discernible pattern. The results highlight the influence of task difficulty on three of the four hypothesised sources of self-efficacy. The degree of influence of task difficulty on each source is outlined in the following sections.

Table 11. Means, standard deviations and correlations for self-efficacy items

Item	M	SD	ME-1	ME-2	ME-3	ME-4	ME-5	ME-6	VE-1	VE-2	VE-3	VE-4	VE-5	VE-6	SP-1	SP-2	SP-3	SP-4	SP-5	SP-6	PH-1	PH-2	PH-3	PH-4	PH-5	PH-6
ME-1	5.4	2.3	**0.96***																							
ME-2	5.6	2.4	0.96	**0.97**																						
ME-3	6.6	2.4	0.80	0.81	**0.87**																					
ME-4	5.3	2.2	0.83	0.84	0.76	**0.91**																				
ME-5	5.5	2.2	0.94	0.95	0.77	0.87	**0.97**																			
ME-6	5.0	2.1	0.90	0.91	0.75	0.85	0.94	**0.95**																		
VE-1	5.5	2.1	0.70	0.70	0.62	0.57	0.68	0.65	**0.92**																	
VE-2	3.7	2.1	0.66	0.64	0.56	0.63	0.66	0.58	0.71	**0.86**																
VE-3	5.6	2.4	0.74	0.72	0.68	0.68	0.73	0.68	0.91	0.76	**0.95**															
VE-4	3.6	2.0	0.75	0.72	0.64	0.63	0.71	0.65	0.68	0.85	0.74	**0.87**														
VE-5	5.1	2.4	0.79	0.77	0.67	0.63	0.75	0.68	0.87	0.73	0.87	0.78	**0.93**													
VE-6	5.9	2.6	0.78	0.75	0.68	0.69	0.75	0.73	0.83	0.69	0.88	0.72	0.83	**0.91**												
SP-1	4.8	2.4	0.80	0.80	0.78	0.72	0.81	0.80	0.65	0.54	0.70	0.62	0.69	0.78	**0.95**											
SP-2	4.9	2.5	0.76	0.76	0.74	0.72	0.78	0.79	0.56	0.50	0.63	0.57	0.61	0.71	0.94	**0.96**										
SP-3	3.0	2.5	0.28	0.32	0.39	0.31	0.30	0.31	0.10	0.16	0.20	0.23	0.23	0.30	0.47	0.57	**0.62**									
SP-4	4.8	2.6	0.73	0.72	0.71	0.69	0.74	0.74	0.56	0.47	0.63	0.53	0.58	0.67	0.91	0.94	0.59	**0.97**								
SP-5	4.5	2.8	0.80	0.81	0.75	0.74	0.80	0.80	0.62	0.55	0.68	0.62	0.64	0.66	0.89	0.87	0.37	0.87	**0.94**							
SP-6	4.5	2.9	0.77	0.78	0.72	0.70	0.76	0.77	0.63	0.55	0.69	0.59	0.64	0.66	0.89	0.86	0.37	0.88	0.98	**0.94**						
PH-1	3.7	1.8	-0.63	-0.59	-0.68	-0.63	-0.60	-0.63	-0.58	-0.42	-0.67	-0.48	-0.57	-0.60	-0.68	-0.69	-0.32	-0.67	-0.64	-0.66	**0.95**					
PH-2	2.4	1.4	-0.27	-0.21	-0.38	-0.34	-0.22	-0.24	-0.26	-0.21	-0.36	-0.22	-0.27	-0.24	-0.39	-0.42	-0.21	-0.40	-0.38	-0.41	0.58	**0.7**				
PH-3	3.6	2.1	-0.58	-0.52	-0.59	-0.59	-0.55	-0.55	-0.58	-0.42	-0.66	-0.43	-0.57	-0.60	-0.65	-0.66	-0.31	-0.67	-0.60	-0.63	0.88	0.58	**0.94**			
PH-4	3.4	2.0	-0.60	-0.57	-0.63	-0.62	-0.59	-0.58	-0.54	-0.42	-0.65	-0.44	-0.54	-0.56	-0.64	-0.65	-0.24	-0.65	-0.65	-0.67	0.90	0.61	0.89	**0.95**		
PH-5	1.8	1.0	-0.41	-0.38	-0.42	-0.37	-0.38	-0.37	-0.35	-0.26	-0.38	-0.19	-0.33	-0.34	-0.40	-0.42	-0.24	-0.37	-0.32	-0.35	0.62	0.33	0.57	0.53	**0.66**	
PH-6	3.4	2.0	-0.51	-0.48	-0.55	-0.52	-0.48	-0.51	-0.50	-0.39	-0.57	-0.38	-0.46	-0.51	-0.55	-0.53	-0.18	-0.55	-0.54	-0.56	0.84	0.56	0.86	0.88	0.53	**0.93**

*Figures in bold diagonal indicate correlation of individual items with overall subscale score

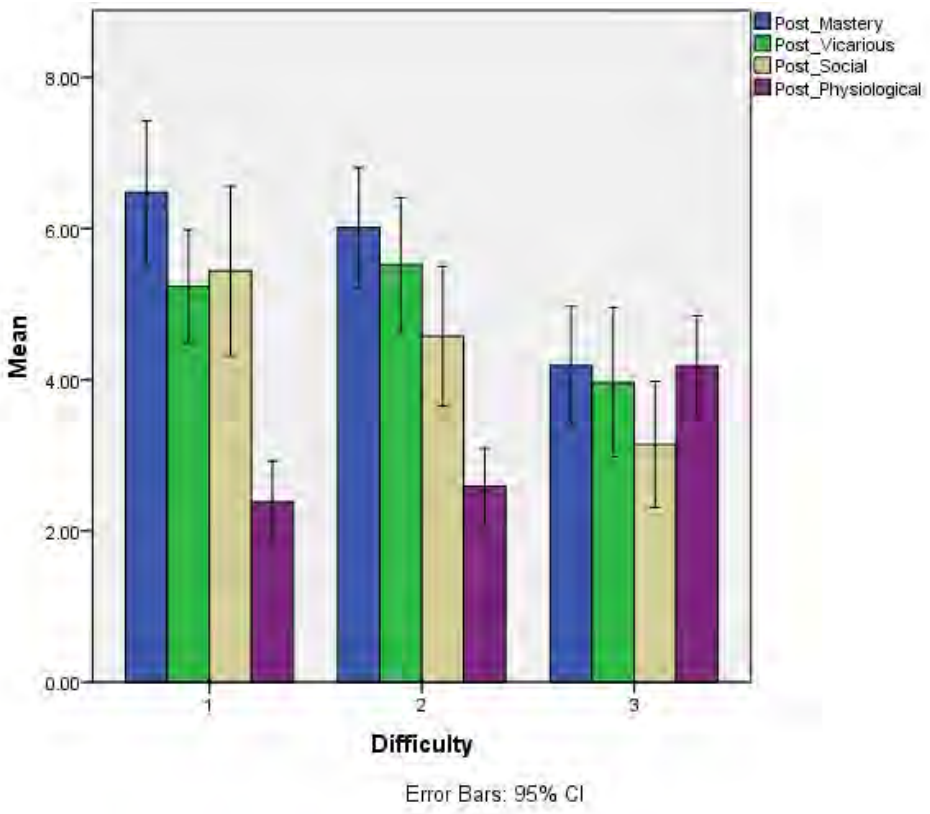

Figure 19. Post-Test Sources of Self-Efficacy Scale Results

The between group ANOVA results outlined in Table 12 highlight statistically significant differences in three of the four sources between difficulty groups. The nature of variance in each source is explored further below.

Table 12. Post-Test Sources of Self-Efficacy Scale Results

	Group	Difficulty	Mean	σ	F	Sig.
Mastery Experience	5A	Low	6.477	2.13	8.911	.001
	5B	Medium	6.015	1.79		
	5C	High	4.189	1.76		
Vicarious Experience	5A	Low	5.235	1.69	3.829	.027
	5B	Medium	5.523	1.98		
	5C	High	3.969	2.23		
Social Persuasions	5A	Low	5.439	2.53	6.204	.003
	5B	Medium	4.576	2.08		
	5C	High	3.144	1.88		
Physiological	5A	Low	2.386	1.21	12.638	.001

State	5B	Medium	2.591	1.14
	5C	High	4.182	1.49

Mastery

Prior to analysis of post-test scale data, an ANOVA was used in order to assess the comparability of groups before exposure to the manipulated difficulty level. No statistically significant difference in pre-test mastery values was observed F(2,63) = 0.392, p =.677. A mixed between-within (repeated measures) ANOVA was employed in order to assess the impact of practice difficulty (High/Medium/Low) on participants' change in mastery values between pre and post-tests. The results suggest a statistically significant difference in mastery value over time between groups, Wilks-Lambda = .889, F (2,63) = 3.922, p = .025, eta squared = .111.

Post-hoc comparisons using the Tukey HSD test suggested that there was a statistically significantly difference between the Low Difficulty group (5A: M = 6.477, SD = 2.13) and High Difficulty group (5C: M = 4.189, SD = 1.76). The Medium Difficulty group (5B: M = 6.015, SD = 1.79) was found to be statistically significantly different to the High Difficulty group but no significant difference was observed when compared to the Low Difficulty group. This is reflected in Figure 20 where the considerable difference between the Low and High Difficulty group means is clear. It is also worth noting the position of the Medium Difficulty group mean. This could indicate that the different difficulty level's influence is not proportionate or the variance could be a product of the relatively small sample size. A much smaller difference in mastery mean value between the Low Difficulty and Medium Difficulty is clear when compared to the difference between High and Medium Difficulty groups.

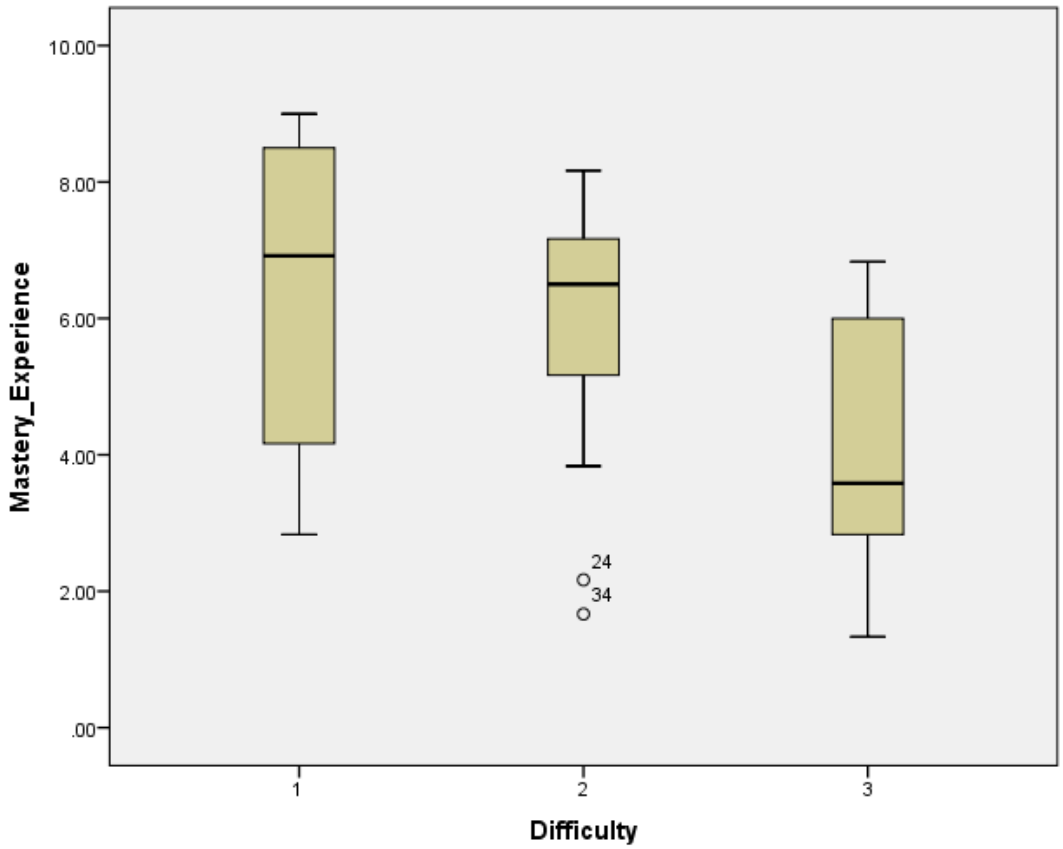

Figure 20. Means and Distributions of the Mastery Source Scale

The results outlined within this section highlight the influence of task difficulty on mastery values. The statistically significant difference between the low difficulty (5A) and high difficulty (5C) suggests a considerable influence. The mean of medium difficulty group (5B) is closer to the mean of low difficulty group (5A) than that of the high difficulty group (5C) (see Table 12). This variance could indicate that a proportionate change in difficulty levels does not result in a proportionate influence on the mastery source. This variance could also be due to the relatively small sample size.

Vicarious Experience

As noted in Table 12 a statistically significant difference in vicarious experience values was observed (p =.027). This, however, did not follow a pattern similar to that observed in the other 3 sources. Prior to analysis of post-test scale data, an ANOVA

was used in order to assess the comparability of groups before exposure to the manipulated difficulty level. No statistically significant difference in pre-test vicarious experience values was observed F(2,63) = 3.093, p =.052. A mixed between-within (repeated measures) ANOVA was employed in order to assess the impact of practice difficulty (High/Medium/Low) on participants change in vicarious values between pre and post-tests. The results suggest a statistically significant difference in improvement between groups, Wilks-Lambda = .846, F (2,63) = 5.722, p = .005, eta squared = .154. Post-hoc comparisons using the Tukey HSD test suggested that there was a statistically significantly difference between the High Difficulty (5C: M = 3.969, SD = 2.23) and Medium difficulty group (5B: M = 5.523, SD = 1.98) (p = .031) groups for the vicarious experience values. This was the only statistically significant difference between groups for the vicarious experience source.

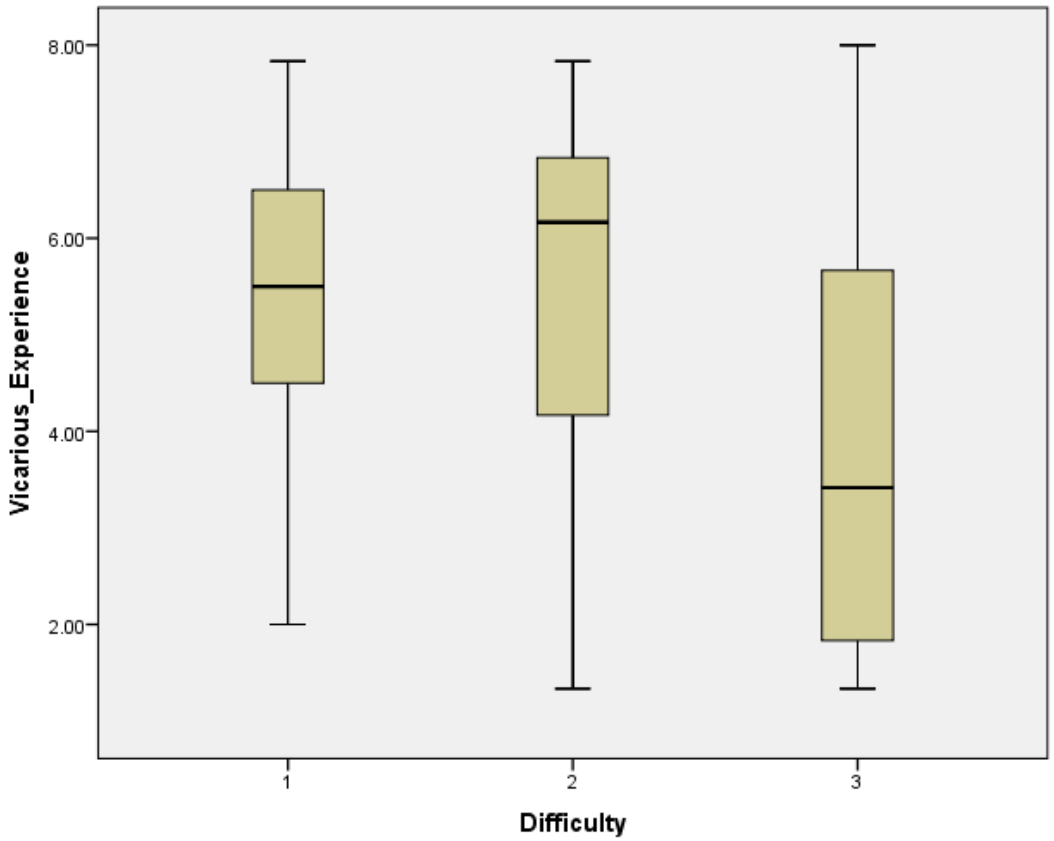

Figure 21. Means and Distributions of the Vicarious Source Scale

The results of the vicarious source subscale suggest that the influence of task difficulty is not as linear as the previously observed influence on the mastery source. As shown in Figure 21 the mean of the medium difficulty group (5B) is higher than that of the low difficulty group (5A) supporting the observed non-linear influence of task difficulty on vicarious experience. Alternatively, the above variance could be a product of the relatively small sample size.

Social Persuasions

Prior to analysis of post-test scale data, an ANOVA was used in order to assess the comparability of groups before exposure to the manipulated difficulty level. No statistically significant difference in pre-test social persuasion values was observed $F(2,63) = 0.031$, $p = .97$. A mixed between-within (repeated measures) ANOVA was employed in order to assess the impact of practice difficulty (High/Medium/Low) on

participants change in vicarious values between pre and post-tests. The results suggest a statistically significant difference in improvement between groups, Wilks-Lambda = .895, F (2,63) = 3.702, p = .03, eta squared = .105.Post-hoc comparisons using the Tukey HSD test suggested that there was a statistically significant difference (p = .003) between the high difficulty (5C: M = 3.144, SD = 1.88) and low difficulty (5A: M = 5.439, SD = 2.53) groups for the vicarious experience values. A statistically significant difference was not observed between the medium difficulty group (5B: M = 4.576, SD = 2.08) and either of the other two difficulty groups.

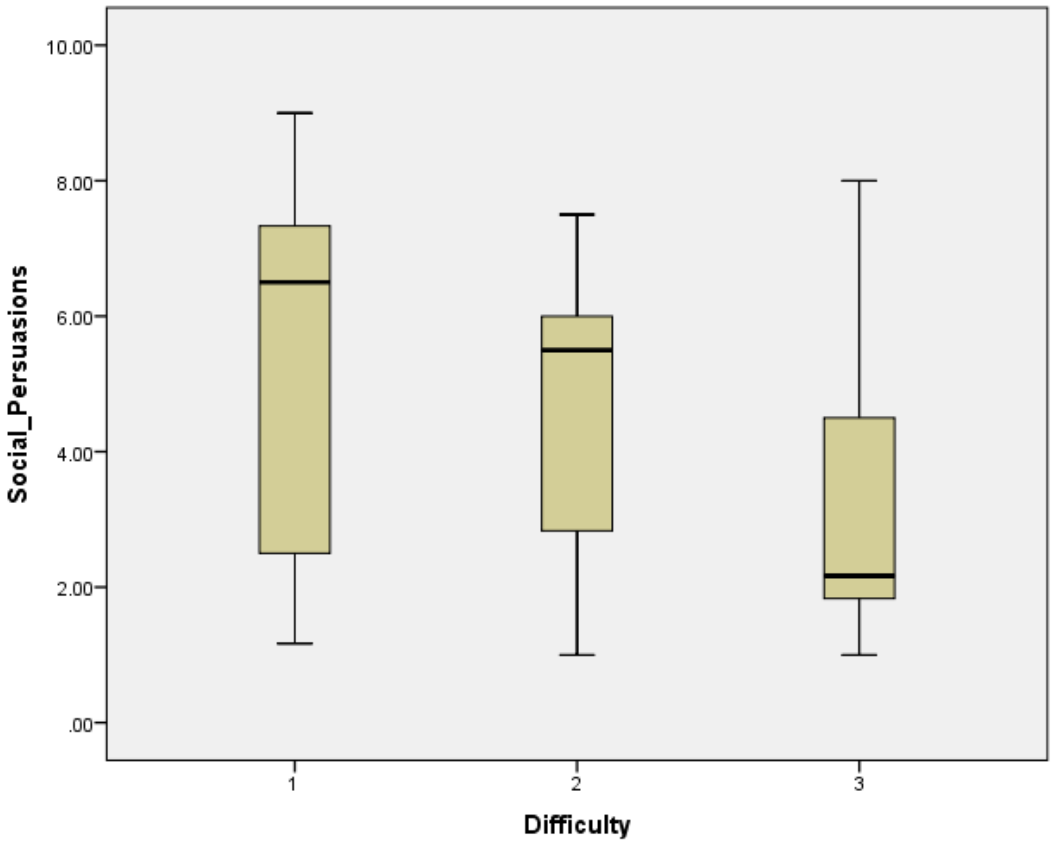

Figure 22. Means and Distributions of the Social Source Scale

The results of the social persuasions subscale present a similar pattern to those observed in the mastery subscale. Task difficulty demonstrates a reasonably linear influence on social persuasions values with the low difficulty group (5A)

demonstrating the highest values while the high difficulty group (5C) demonstrated the lowest social persuasions reports. These results suggest that task difficulty influences the vicarious experience source of self-efficacy.

Physiological State

Prior to analysis of post-test scale data, an ANOVA was used in order to assess the comparability of groups before exposure to the manipulated difficulty level. No statistically significant difference in pre-test physiological state values was observed $F(2,63) = 0.197$, $p = .822$. A mixed between-within (repeated measures) ANOVA was employed in order to assess the impact of practice difficulty (High/Medium/Low) on participants change in vicarious values between pre and post-tests. The results suggest a statistically significant difference in improvement between groups, Wilks-Lambda = .865, F (2,63) = 4.913, p = .01, eta squared = .135. Post-hoc comparisons using the Tukey HSD test suggested that there was a statistically significantly difference between the High Difficulty (5C: M = 4.182, SD = 1.49) and Low Difficulty (5A: M = 2.386, SD = 1.21) groups for the vicarious experience values. A statistically significant difference was not observed between the Medium difficulty group (5B: M = 2.591, SD = 1.14) and either of the other two difficulty groups.

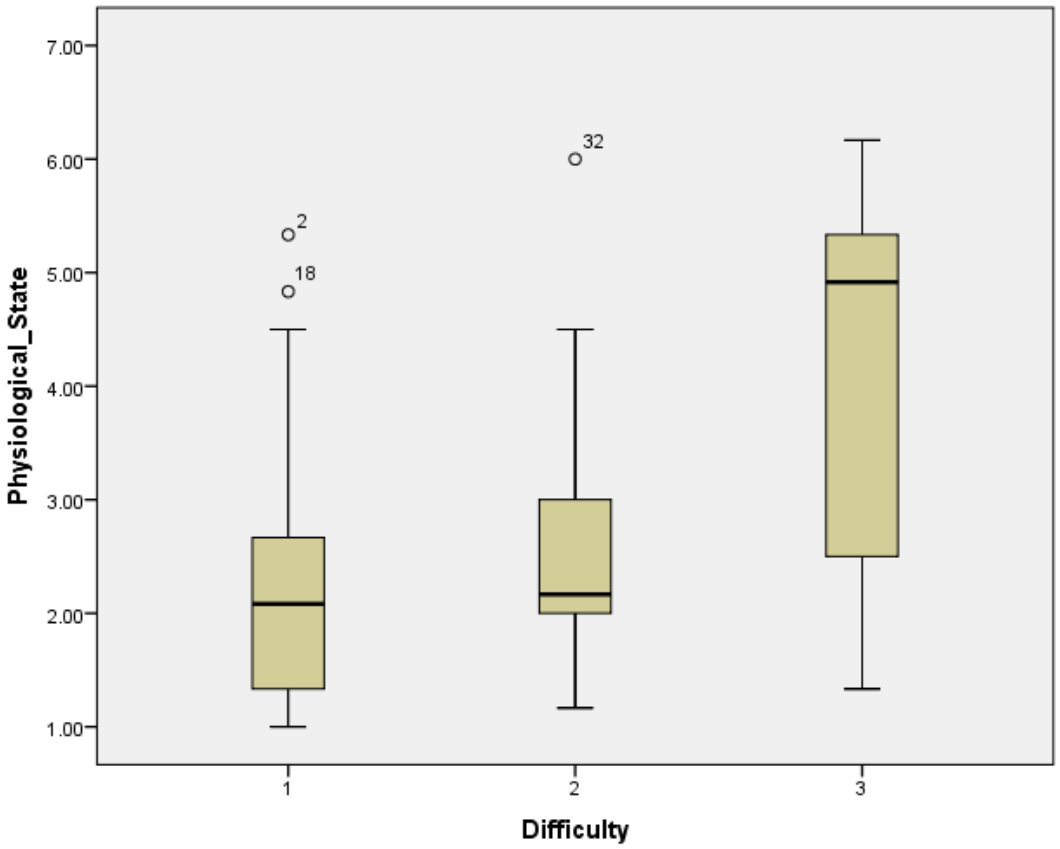

Figure 23.Means and Distributions of the Physiological Source Scale

As previously discussed the physiological source subscale uses negative response items. This is in contrast to the three other source subscales. Items within this subscale focus on anxiety or stress physiological manifestations. The negative report nature of this subscale results in an opposite slope pattern when compared to the previous three sources. The results indicate that those who practiced with the high difficulty version of the task (group 5C) reported markedly higher physiological source scores. This indicates higher anxiety/stress reports. Those who practiced with the low difficulty (group 5A) reported considerably lower scores indicative of lower anxiety/stress associated with the task. The results of this source suggest that task difficulty exerts a considerable influence on physiological state.

Results of Scale Relative to Performance and Engagement

The relationship between the sources of self-efficacy and performance are outlined in Table 13. Pearson's correlation results are shown between the sources and the improvement in performance scores. Improvement in performance scores was calculated as an expression of how much the post-test score exceeded the pre-test.

Table 13. Relationship Between Improvement and Self-efficacy Sources

	Improvement	Mastery Experience	Vicarious Experience	Social Persuasions	Physiological State
Percentage Improvement	1				
Mastery Experience	.708**	1			
Vicarious Experience	.557**	.805**	1		
Social Persuasions	.643**	.820**	.683**	1	
Physiological State	-.508**	-.623**	-.588**	-.660**	1

**. Correlation is significant at the 0.01 level (2-tailed). N = 66

Percentage improvement demonstrates the largest correlation with mastery experience (r = .708, p<.01), but also demonstrates considerable correlations with vicarious experience (r = .557, p<.01) and social persuasions (r = .643, p<.01).

Similar data examining the relationship between the sources of self-efficacy and engagement rates is outlined in Table 13. Pearson's correlation results are shown between the sources and engagement rates. Engagement rates are calculated by the amount of attempts participants made using the practice task throughout the defined practice period.

Table 14. Relationship Between Engagement Rate and Self-efficacy Sources

	Engagement Rate	Mastery Experience	Vicarious Experience	Social Persuasions	Physiological State
Engagement Rate	1				
Mastery Experience	.741**	1			
Vicarious Experience	.577**	.805**	1		
Social Persuasions	.693**	.820**	.683**	1	
Physiological State	-.517**	-.623**	-.588**	-.660**	1

** Correlation is significant at the 0.01 level (2-tailed). N = 66

The Sources of Self-Efficacy Scale data demonstrates considerable links between the sources of self-efficacy, performance and engagement rates. A Pearson's correlation test examining the relationship between practice attempts and improvement, using all five rounds of data, resulted in a large positive correlation r = .525, n = 317, p = .001. This indicates a shared variance of 27.6% (Cohen 1988, p.79-81). An additional Pearson's test using only round 5 engagement and improvement resulted in a larger correlation (r=0.734, n=66, p<0.001). This suggests that raised engagement and performance levels are indicative of raised self-efficacy levels while conversely lowered engagement and performance levels are indicative low self-efficacy levels. Previous rounds that manipulated task difficulty noted similar links to engagement and performance. The engagement rates, self-efficacy source values and performance figures for the high difficulty (5C) and low difficulty (5A) groups from round 5 are presented graphically in Figure 24 below.

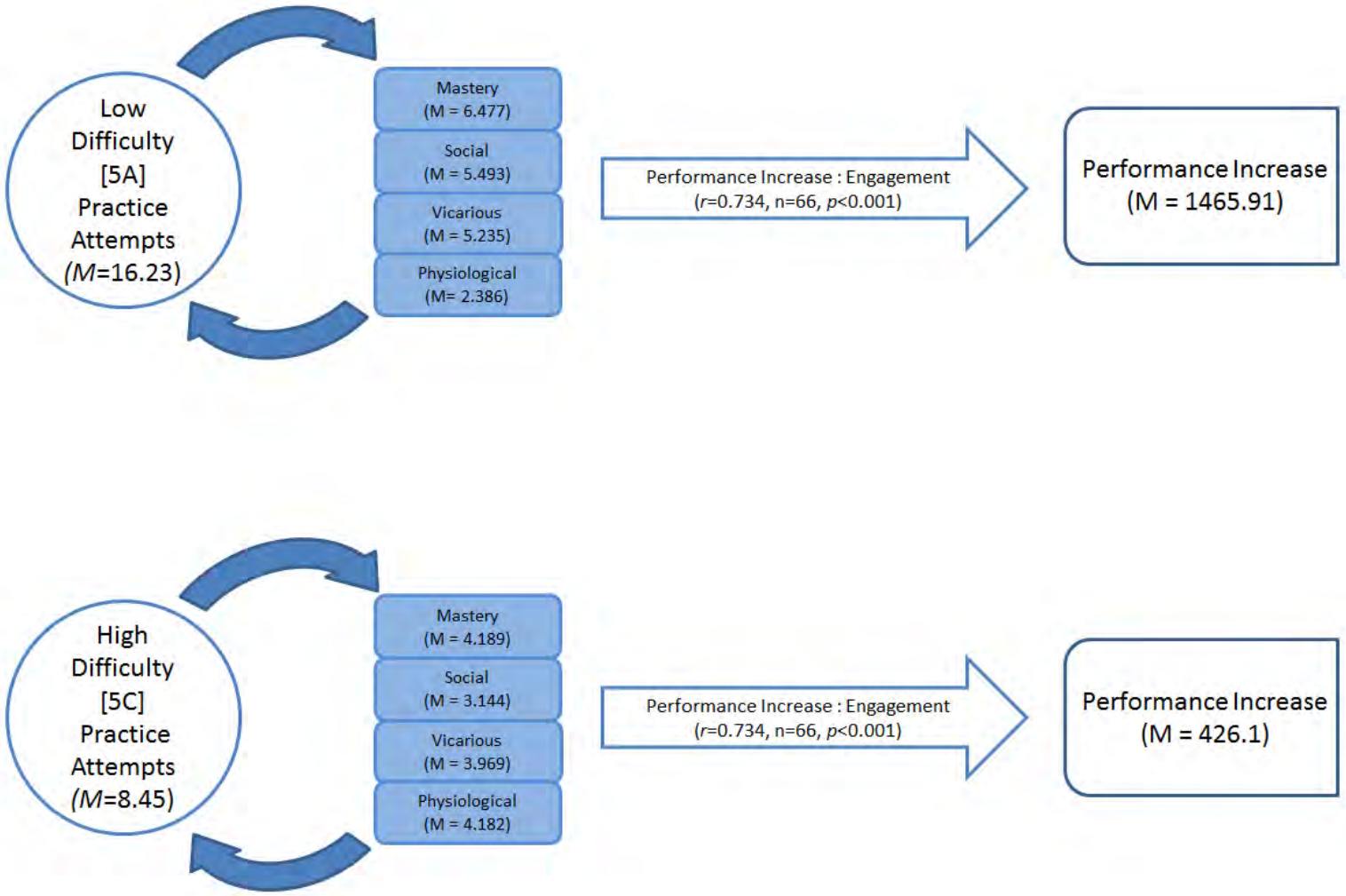

Figure 24. Reciprocal Model of Performance and Self-efficacy Formation

Limitations

Treiman (2014) emphasises the importance of the researcher being aware of the inherent limitations of the data when analysing and interpreting results. The sample size of each round of data limits the power of the statistical results and also the scope of methods available for analysis. The considerable standard deviation and mean variance is, at least in part, due to this relatively small sample size. This limits the nature of conclusions drawn. It also limits the statistical models that can be employed. For example factor analysis methods are unsuitable when examining the scale data due to the sample size not meeting minimum guidelines (Treiman 2014). The practical limitations of sample size arose from the school environment in which the study took place. In order to allow for two rounds of testing within one school year, a round could not exceed half the participants of the total population of that year. Due to the multiple rounds of testing two rounds of testing per school year was necessary in order to keep the project within reasonable timeframe constraints. This limitation could have been circumvented if additional schools were recruited to take part in the study, however, this raises further problems related to comparability and shared environment. This would, however, require greater resources and improvement of the computerised maze navigation task.

Linked to the sample size is the generalisability of the findings. In order to work towards generalisable findings this study employed randomised selection and randomised grouping. Randomised selection ensured that participants from all class groups were present in each round. In addition, participants were then randomly assigned to one of three groups within each round. While this encourages comparability within the total sample, generalisability cannot be assumed for the national population or beyond the school context. However, the findings of Lynch *et*

al. (2013) suggest that the performance within the task is comparable with similar age groups within a considerably more homogeneous school environment. Further applications of a similar method employed within this study in disparate environments would aid in the examination of the generalisability of the findings of this study. A novel approach that could potentially provide a much larger and diverse sample size would be the development of the maze navigation task into a cross platform format. With an automated data collection system, considerable user data and the potential of a continuous variable adjustment the potential resulting data could provide exceptional opportunities for theory development. Finally, scale alterations should be considered in terms of validity and reliability. Neither are assumed requiring further investigation post data collection. For a more complete evaluation of validity of the scale used see the Data Analysis subsection within chapter three.

CHAPTER FIVE: DISCUSSION

The discussion chapter follows three major topics arising from the results of this research project. By examining these topics in the context of current theory, a broad theoretical base for the subsequent conclusion chapter is established. This allows for a comprehensive examination of the results of the current research project, from which the research questions that underpinned this research project are explicitly addressed in the concluding chapter. The discussion begins with an outline of the influence of task difficulty. This influence is examined using multiple rounds of data and the stability of this influence is also considered. This impact is strongly supported by statistically significant round 5 results of a mixed between-within (repeated measures) ANOVA (See p.95) examining performance increases across difficulty groups. Additional mixed between-within (repeated measures) ANOVAs examining round 1 (See p.75) and round 2 (See p.79) data, although marginally statistically insignificant, echo this influence. When examined in the context of limited sample sizes, and consistent effect sizes, these marginally statistically insignificant results appear to be trending towards significance. Within this topic the role of reward is also considered. As difficulty and reward have the potential to affect an individual's perception of the outcome of a task (Gilbert *et al.* 2012, Fröber and Dreisbach 2014) both are examined in terms of their comparative degree of influence.

The second topic builds upon the influence of difficulty and examines its impact on the sources of self-efficacy. This impact is first examined in terms of outcomes that have been previously linked to self-efficacy, such as engagement and performance (Komarraju and Nadler 2013, Parker *et al.* 2014). This impact is further examined using the sources of self-efficacy data from round 5. This impact is highlighted by

the statistically significant results of mixed between-within ANOVAs examining the change of source values between difficulty groups. The conclusion of this topic examines the theoretically reciprocal relationship between performance and self-efficacy (Bandura 1997). The third and final topic examines variances in engagement and performance data as potential indicators of self-efficacy manipulation. The relationships between engagement, performance and sources of self-efficacy data recorded in round 5 provide an empirical comparison for the previously discussed theoretical foundation.

The Influence of Difficulty

Rounds 1 and 2 were designed to examine the influence of task difficulty on engagement and performance. Examining the results within the context of self-efficacy theory, the data from round 1 and 2 indicates that lower difficulty levels lead to increased engagement. This increased engagement allowed for more frequent opportunities to experience mastery, which in turn can further increase self-efficacy (Liem *et al.* 2008). This is suggestive of a reciprocal relationship between self-efficacy and performance as hypothesised by Bandura (1994). This hypothesised reciprocal interaction is expanded upon further later in this discussion. The increased engagement levels observed in participants practicing with the lower difficulty versions of the task should be noted in the context of consistently large correlation values between engagement and overall performance observed within rounds ($.734 \geq r \geq .494$). Comparisons of data from all rounds consistently demonstrate a considerable negative influence of difficulty on engagement. In addition data from round 5 suggests that lower task difficulty led to increased self-efficacy. The nature of the formation of self-efficacy is discussed in parallel to round 5 data later in this chapter. The manner in which self-efficacy changes due to the reciprocal relationship

between self-efficacy and performance, as hypothesised by Bandura (2012), is also examined in the later stages of this discussion.

The influence of difficulty on engagement, and consequently performance, is well supported when examined from a resultant perspective (Orvis *et al.* 2008, Kumar and Jagacinski 2011, Rosander and Eriksson 2012, Fraser *et al.* 2014). However, the manner in which difficulty influences engagement is less clear. The reciprocal model explaining the relationship between self-efficacy and performance proposed by Bandura (2012) helps explain the links between engagement, performance and sources of self-efficacy data observed in this study. The empirical support for the reciprocal relationship between self-efficacy and performance is based on the widely agreed link between increased opportunities to experience mastery and resultant increased self-efficacy (Bandura 1977, Schunk and Pajares 2009, Usher and Pajares 2009). This in turn leads to increased engagement, which allowed for further mastery opportunities. Rather than increased self-efficacy in isolation leading to increased performance, this suggests that increased self-efficacy led to increased engagement. This ultimately resulted in increased performance. This is supported by the results of an ANCOVA which highlighted a non-significant difference in performance increases when engagement is controlled for (See p.64). Previous studies have suggested that increased self-efficacy leads to increased engagement rates and retention in longitudinal tasks (Caraway *et al.* 2003, Annetta *et al.* 2009, Bresó *et al.* 2011); while others have focused on self-efficacy's performance enhancement (Williams and Williams 2010, Sitzmann and Yeo 2013). However, no study identified in the literature review that informed this study observed changes in behaviour that were attributed to varied self-efficacy levels. This highlights a limited interpretation of self-efficacy measure relative to a once off task or performance in a

task with a prescribed practice period. Self-efficacy is broadly identified as a self-regulation theory (Sitzmann and Yeo 2013). Studies that do not allow for observation of self-regulated action fail to examine performance enhancing self-efficacy effects that stem from autonomous behaviour alteration, such as the observed engagement rates in this study. Although self-efficacy levels have demonstrated predictive value for performance in one off tasks, this type of experiment only allows for a shallow examination of the impact of self-efficacy on performance (Bandura 2012).

In summary, the negative influence of difficulty on engagement, and as a result performance, is supported empirically by the results of multiple rounds of data within this study. In addition, there exists considerable theoretical support within self-efficacy theory for this influence (Zimmerman 2000, Schunk and Pajares 2009). This influence aligns with the proposed reciprocal relationship between performance and self-efficacy formation as proposed by Bandura (2012). However, the manner in which difficulty influences each source of self-efficacy and the nature of the relationship between self-efficacy and performance remains unclear. These points will be further explored in subsequent sections of this discussion by examining round 5 sources of self-efficacy data alongside performance data. Prior to an examination of these points, the consistency of the influence of task difficulty is first explored.

Stability of Difficulty Influence

The stability of the previously identified difficulty influence can be examined using data from rounds 1, 2 and 5. The data from these rounds support the conclusion that hidden difficulty variations, combined with increased opportunities for comparisons, had a negative impact on engagement rates and performance that are theoretically

indicative of self-efficacy (Liem *et al.* 2008, Bresó *et al.* 2011). Comparisons between high difficulty groups in round 1 and 2 are particularly interesting. The consistency of the influence of difficulty is evident across multiple rounds. Within round comparisons of difficulty groups for round 1 data shows a considerable difference in overall performance and engagement (see Table 4, p.76). The same trend is evident to a lesser degree in round 2 (see Table 5, p.80) and round 5 (see Table 8, p.96). While the influence of varying difficulty within rounds is clear, there is a considerable difference in overall performance and engagement levels when round 1 data is compared to round 2 data. This variation is attributed to the difference in practice environment with round 1 practicing the task away from the typical school environment while round 2 practiced the task over a typical school week. Theoretically the most direct influence of the altered environment would act on the social persuasions source (Bandura 1986). This is based on the increased opportunities for peer comparisons within the school environment.

This source draws on the contribution of peers, and authority figures, in the formation of self-efficacy (Bandura 1986). This is especially relevant to the current study where participants are presented with a novel task, with which they have no prior experience to aid self-appraisal of performance. As highlighted by Usher and Pajares (2008) when individuals "are not yet skilled at making accurate self-appraisals, students often depend on others to provide evaluative feedback and judgments about their performance" (Usher and Pajares 2008, p.754). However, it is important to note that social persuasions are not always positive in nature. As highlighted by Bandura (1997) negative social persuasions have the potential to be extremely damaging to self-efficacy formation.

While it was not possible to examine the level of peer comparisons within this study, the considerably lower engagement rates and overall performance observed in the in-school (round 2) participants suggest that the posited positive impact of comparisons is questionable. This highlights an emerging trend of self-efficacy research that examines the social and peer influence (van Dinther *et al.* 2011) on self-efficacy formation. The comparison of round 1 and round 2 data raises questions relating to the primarily positive view of the influence of social comparisons (Lam *et al.* 2004). There exists theoretical support for social comparisons being primarily negative in nature (Bandura 1997). This is especially true where participants are unaware of factors impacting their performance such as the hidden difficulty variations used in this study. This supports the findings of Usher (2009) the potential negative impact of these comparisons. The negative nature of these comparisons arises from the need to gauge one's performance primarily through socially defined standards.

> "The problem of performance ambiguity arises when aspects of one's performances are not personally observable or when the level of accomplishment is socially judged by ill defined criteria so that one has to rely on others to find how one is doing. In the latter situations, if designating feedback is lacking for tasks on which performers cannot judge their output, they are left in foggy ambiguity . . . in most everyday pursuits, such problems do not arise because people have aims in mind, and they do not need others to tell them their performances because they can see for themselves how they are doing"
>
> (Bandura 1986, p. 398)

Noting that raised engagement and performance are indicative of raised self-efficacy (Liem *et al.* 2008, Bresó *et al.* 2011), the data from rounds 1 and 2 suggests that the school environment was not conducive to self-efficacy development. As highlighted by Bandura (1997) social influences can exert considerable negative influences on self-efficacy if the individual perceives themselves to be underperforming relative to

their peer groups. When considered in terms of impact on the sources of self-efficacy, the link to social persuasions and vicarious experience is evident (Bandura 1997). However, it is worth considering the negative impact of the school environment in terms of the most influential source, mastery.

When individuals engage in a new task or activity they will ascertain success or failure through comparisons to peers (Usher and Pajares 2008). This has considerable negative connotations for those unknowingly practicing with the high difficulty task. Usher and Pajares (2008) suggest that perceived inferior performance relative to a comparable peer can have considerable negative impacts on self-efficacy. The hidden manipulation of difficulty within this study amplifies this influence. Participants' perceptions of success would be considerably negatively skewed when comparing themselves to peers practicing with what they believe is an equal task but is, in fact, a considerably easier version. This is especially relevant given that "it is often easier to diminish a student's self-efficacy with negative social persuasions than to enhance it with positive messages" (Britner and Pajares 2006, p.495).

The previously mentioned studies suggest that those practicing with the highest difficulty versions would be subject to the greatest negative affects of peer comparisons. This suggests that a higher rate of comparisons would lead to further negative influences on self-efficacy formation (Usher 2009). Indicative outcomes such as performance and engagement should, therefore, be lower in those who engaged in increased comparisons. This supports the interpretation that lower engagement and performance observed in those practicing throughout the course of a typical school week was due to the negative influence of peer comparisons. The deviation between groups in the in-school round 2 data was also higher with a

comparatively low overall mean when compared to round 1 data. This suggests that without opportunities for negative comparisons those practicing with the high difficulty version of the task, in comparative isolation, did not suffer from earlier negative evaluations of mastery to the same extent. Ultimately this resulted in a greater level of engagement and associated increases in performance. This once again highlights the importance of perception in the formation of self-efficacy.

Noting the lack of data relating to frequency and quality of social comparisons, it is important to consider these points in light of the studies limitations. The nature of interaction within each environment went beyond the remit of the study. Given the advent of constant connection through social media platforms, the assumption that the school environment facilitated increased peer comparisons could be flawed. A myriad of further unidentified variables could also have influenced engagement; such as a lower competing workload for those practicing in the out of school environment. While the nature of the role that the practice environment plays remains unclear, the consistent impact of task difficulty within both settings is evident. Bandura (2006) describes the contribution of each source as informative in nature meaning that individuals draw information from each source in order to form a sense of efficacy. From this theoretical basis, varying task difficulty is posited to most directly influence the mastery source. The data suggests that increased difficulty led to lower levels of mastery experience, while conversely lower levels of difficulty led to higher levels of mastery experience. The manner in which individuals interpret success in the given task remains unclear, however results from round 4 suggest that the inbuilt scoring system was not the primary manner in which participants estimated their own capability. This supports the earlier suggestion that peer comparisons are a key factor when interpreting mastery source information.

The Role of Reward

Round 3 was designed to address the question of how success was interpreted; the results from this round suggest an interaction effect between difficulty and reward leading to overall scores that exceeded all rounds where difficulty or reward were manipulated in isolation. In order to examine whether the performance observed in the previous two rounds was ascertained based solely on the score output displayed to each user as they completed the task, the scoring system was manipulated. The inflation of this scoring system was designed to negate the negative impacts of increased difficulty levels. This raises the question of whether the scoring system is viewed as a feedback mechanism, as a reward mechanism or potentially both. In order to examine these areas in greater detail the task was altered to create a proportional reward system. This resulted in three versions of the task: High Difficulty/High Reward, Medium Difficulty/Medium Reward and Low Difficulty/Low Reward. The rationale behind this system is based on the positive effects of the increased reward negating the negative effects of the increased difficulty. The degree of negation provides an indication of the degree to which participants rely on the score output in order to judge their own success or failure in the given task. Bandura (1986) outlines how rewards can act as a source of information as well as motivation. The information aids in the individuals' assessment of their own performance, while anticipation of a desirable outcome motivates the individual. Schunk and Zimmerman (2012) outlined performance-contingent rewards as being especially conducive to self-efficacy development due to a combination of the informative and motivational properties, similar to those outlined by Bandura (1986).

Based on the previously outlined benefits of increased rewards, the predicted impact of this proportional reward system was a negation of the adverse effects of increased difficulty. This would manifest with groups' post-test scores returning to an overall mean as the higher reward positively influenced the underperforming high difficulty group and the low reward system negatively impacted the previously over performing low difficulty group. The impact of the reward variation could be estimated by the degree to which group means were altered by the increased or decreased reward setting. However, the results of round 3 did not demonstrate this pattern. Rather, a considerable overall increase in improvement of all groups within round 3 was observed. The data from the round demonstrated that increasing reward did not simply counteract the negative influence of difficulty on engagement and performance. The overall elevation of engagement and performance suggested a potential interaction effect between the difficulty manipulation and reward manipulation. In order to explore the nature and impact of this interaction a further round of testing was devised.

Round 4 focused on investigating the potential interaction effect observed in the previous round. In order to examine whether an interaction affect had occurred, and the impact of reward without difficulty manipulation, this round employed a version of the task that used the reward variations outlined in round 3 without varying difficulty levels. The isolated impact of the varying reward levels led to relatively minor variations in post-test scores between groups (See Figure 14, p.94). Round 2 provided data that allowed for an examination of the influence of difficulty manipulation while round 4 allowed for an examination of the influence of reward manipulation. However, the variance observed in round 3, attributed to the

proportional reward setting, exceeded that observed in round 2 and 4. This supports earlier suspicions of an interaction effect.

Social models have been shown to be particularly important when an individual has little prior experience in a task. This is supported by the earlier work of Schunk and Hanson (1985) who focused on the impact of peer models on children's self-efficacy and achievement. Comparisons where the individual perceives themselves as negative in what a peer perceives to be an easy task can have disproportionately negative impacts (Blanton *et al.* 1999). This is especially relevant to the current study given the manipulation of difficulty and noting the overall improved engagement and performance rates that are attributed to the interaction effect between reward and difficulty. Similar research suggests that transitional periods, such as the transition between elementary and middle school, comparable to that which the participants in the current study were experiencing, enhances the impact of social comparisons (Eccles *et al.* 1984). These comparisons and susceptibility to social influence align theoretically with the social persuasions source, vicarious experience source and perhaps less intuitively, the mastery source. When criteria for success is unclear, or ill-defined, participants are likely to ascertain whether an outcome was positive or negative (mastery) by comparing themselves to others (Bandura 2012, Zelenak 2015, Honicke and Broadbent 2016).

These theoretical links are supported by the data from round 3 relating to engagement and performance. Overall a considerably higher rate of engagement was noted across all groups within round 3 when compared with any other round of data (see Table 7, p.90). In addition, the differences in group engagement figures was lower than that observed in other rounds. Final performance figures were also greater than those observed in any other round. The considerable correlations between

engagement rates and overall improvement remained comparable to figures observed in other rounds. When examined in light of round 4 data the impact of varying reward levels as the only manipulated variable was considerably less than the proportional reward/difficulty settings used in round 3. This is also evident in round 2 data where difficulty was the only manipulated variable. The relatively minor variance attributed to reward, considered relative to the greater variance observed in round 2 or round 3 data, suggests that participants are assessing their own performance from additional sources besides the score based reward system. This is worth considering in light of Bandura's (2012) assertion that the sources of self-efficacy should be considered informational in nature. While the nature of the information influences each source, and as a result self-efficacy, so too does the manner in which an individual interprets this information (Bandura 1977). The act of interpretation supports the views of Fröber and Dreisbach (2014) who posit that feedback is not separable from reward. As previously discussed with a theoretical focus, the results of rounds 2, 3 and 4 suggest that the proportional reward system allowed for social comparisons that were more conducive to self-efficacy formation. The higher engagement rates and overall performance align with theoretically predicted outcomes of enhanced self-efficacy (Parker *et al.* 2014, Zelenak 2015).

The Impact of Difficulty on Self-Efficacy Formation

The first four rounds of testing focused on task variables that could potentially influence self-efficacy formation over the course of the practice period and found that task difficulty consistently impacted engagement and performance, which have been previously linked to variances in self-efficacy (Bresó *et al.* 2011, Sitzmann and Yeo 2013). Reward was also found to impact engagement rates and performance, albeit to a lesser degree. The impact of these variables align with theoretical

predictions and the between group comparisons, within each round, lend credibility to tentative claims of internal validity. In an effort to examine any potential direct link to self-efficacy theory a scale examining the sources of self-efficacy was employed in round 5 of testing. Prior to discussing the results of this scale, in conjunction with the related performance and engagement figures, it is essential that the limitations of the scale are established. The sources of self-efficacy scale developed by Usher and Pajares (2009) reflects a minority of studies that abide by the theoretical underpinnings of self-efficacy (Bandura 2006). A later review examining the functional properties of self-efficacy confirmed the widespread misinterpretation of elements of self-efficacy theory in the design of studies often resulting in invalid data (Bandura 2012). Although the substantial development study suggested considerable validity (Usher and Pajares 2009), there is always a concern that adaptations can, and often do, invalidate instruments (Arndt and De Bruin 2011). Theoretically there is support for the alteration of an instrument that further focuses the domain of operation (Honicke and Broadbent 2016). Previous studies examining the explanatory and predictive value of self-efficacy suggest it has increased in line with the specificity of the domain of operation (Bandura 1997, Pajares 1997). Conclusions drawn from the use of such an altered scale must remain tentative until further large scale applications and subsequent factor analysis methods are employed. However, apparent internal validity can be used in order to examine whether the scale is suitable for further large scale applications. These inferences are first drawn from theoretical predictions and accompanying performance data. Secondly comparisons of scale data are made between difficulty groups within round 5. Predictions relating to the impact of varying difficulty on the sources of self-efficacy have been previously based on observed performance variance. By

considering the scale results in terms of performance we can examine this deduction from an empirical and theoretical perspective.

A distinct pattern is evident in three of the four sources when comparing between difficulty groups (See Figure 19, p.105). This is further supported by the results of mixed between-within ANOVAs examining change in source values across difficulty groups. Each source demonstrated statistically significant results and are discussed in greater detail, relative to individual sources, below. Participants who practiced with the high difficulty task consistently reported the lowest mastery, social and physiological source values. Participants who practiced with the medium difficulty task reported increased values across the mastery, social and physiological sources, while participants who practiced with the low difficulty task reported the highest values for these three sources. However, this pattern was not observed in the vicarious experience source. While this could be attributed to variance as a result of a small sample size, it is worth noting that the vicarious source is acknowledged as potentially the most difficult source to measure given the considerable amount of confounding factors (Bandura 2012, Honicke and Broadbent 2016). Each source will be discussed independently below.

Impact on Mastery

The impact of varying task difficulty on mastery experience is clear (See Figure 20.) Those who completed the practice period using the low difficulty version of the task reported the highest levels of mastery experience. Conversely, those practicing with the highest difficulty version of the task reported a markedly lower level of mastery experience. In rounds one, two and three the impact of difficulty was theoretically attributed to greater opportunities for mastery experience. It is important to note that increased performance figures are not suggested to represent mastery experience.

Rather the cyclical nature of a perceived positive outcome, leading to increased engagement, facilitates more frequent opportunities for mastery experience to occur. This is presented graphically with accompanying data in Figure 24 (p.115). This link was theoretically supported in terms of the resultant behaviour aligning with predicted outcomes.

A lower level of difficulty results in a greater likelihood that the individual will perceive the outcome to be positive (Joo *et al.* 2013). This is especially true when the participant is not aware that they are using a task with a higher difficulty level than their peers due to participants using peer comparisons to ascertain success or failure in a task that is new to the individual (Liem *et al.* 2008). The additional scale data in round five provides greater clarity with regard to the manner in which the sources of self-efficacy were altered as a result of task difficulty. Noting that mastery is the most influential source of self-efficacy (Bandura 1997), the considerable impact of task difficulty on this source supports previously reasoned theoretical links between mastery experience and difficulty. This should be considered in light of the engagement rates of each difficulty group (See Figure 17) and also in the context of round 5 source values which were examined using a mixed between-within (repeated measures) ANOVA. This highlighted a statistically significant difference in the change in mastery values within difficulty groups between the pre and post-test, Wilks-Lambda = .889, F (2,63) = 3.922, p = .025, eta squared = .111 (See p.106).

However, the manner in which individuals interpreted this information remains somewhat unclear. In round four participants were divided into three groups based on manipulation of the score system. This resulted in comparatively minor differences in performance when groups within round 4 were compared. This suggests that participants were assessing their own performance using additional

information besides that of the scoring system. This is supported by Bandura (1997) who comments on the dynamic nature in which an individual evaluates their own performance. It should also be noted that Bandura (2006) suggests that the sources are interdependent. The influence of social persuasions, for example, could have a significant bearing on what an individual deems a positive outcome. It is plausible, and theoretically supported, to assume that task difficulty is impacting on multiple sources. This point is explored further in the following sections within this chapter.

Impact on Vicarious Experience

The consistent pattern between difficulty groups observed in mastery experience, social persuasions and physiological state is not evident when examining vicarious experience. Although a statistically significant difference between low difficulty and high difficulty groups was observed; the medium difficulty group mean was higher than the low difficulty group mean (See Figure 21.). This is representative of the findings of the majority of studies that have attempted to reliably measure this source (Usher and Pajares 2009). Harris (1995) highlights the complicated nature of this source. She suggests that the roles of the model have a considerable impact on the nature of vicarious experience drawn by the individual. "This is likely why, with few exceptions, researchers have reported low to modest reliability coefficients among items created to assess vicarious experience" (Usher and Pajares 2009, p.90).

Acknowledging the considerable variance in means, the aforementioned statistically significantly different vicarious experience values between those practicing with the high and low difficulty versions of the task are worthy of discussion. The results suggest that the difficulty variable had an impact on the vicarious source although this impact was not as clear, or consistent, as the impact of difficulty on the other three sources. The matrix in Table 11 (p.104) shows comparatively high inter-item

and item to subscale correlations. This suggests reliability and supports the findings of Usher and Pajares (2009) and Harris (1995) who outline the dynamic and nuanced influence of social modelling. Despite the considerable design and validation efforts of Usher and Pajares (2009), it appears as though the vicarious source of self-efficacy was not supported by the data produced by the current research project.

Impact on Social Persuasions

The social persuasions source results demonstrate a similar trend to that seen in mastery and physiological state. The results of a mixed between-within (repeated measure ANOVA indicate that varying difficulty levels resulted in significantly different social persuasion reports Wilks-Lambda = .895, F (2,63) = 3.702, p = .03, eta squared = .105. This consistent pattern shows those who practiced with the lowest difficulty setting reported the highest social persuasion values (M = 5.439, SD = 2.53). Conversely, those that practiced with the highest difficulty setting reported the lowest social persuasions values (M = 3.144, SD = 1.88). This supports the findings of Gainor and Lent (1998) and Brown and Cialdini (2015) who posit that perceived performance can alter an individual's susceptibility to social influence. These results are further supported by multiple studies that noted considerable co-variance between the mastery and social persuasions sources (Usher and Pajares 2008, Schunk and Pajares 2009, Tsai *et al.* 2011).

Bandura's reciprocal determinism (Bandura 1986) is based on the concept that an individual's performance impacts on their environment and self-beliefs. Subsequently, an individual's performance can be influenced by this newly altered environment or self-belief. This interaction between behaviour, cognition and environment is termed triadic reciprocality (Bandura 2012). Examining the data within this context suggests that individuals' perceptions of their own competence

influenced the manner in which they perceived social persuasions source information. Bandura's (1986) triadic reciprocality suggests that the manipulation of difficulty within round 5 of the study impacted the environment and behaviour of the participant. The results of this round support this, demonstrating a considerable difference in social persuasion reports when compared across difficulty groups. This is also reflected in results linked to the behavioural component of the triadic reciprocality model. Data relating to engagement rates and performance indicate a statistically significant difference in performance when compared across difficulty groups (see Table 8, p.96).

Impact on Physiological State

The physiological state subscale demonstrates a negative slope pattern when compared with other sources (see Figure 23). This is due to the negatively weighted nature of the subscale used when examining this source (Usher and Pajares 2009). The items used in this scale focus on physiological cues indicative of negative arousal including anxiety and stress (Usher and Pajares 2009). This is typical of scales designed to examine the physiological state source (Bandura 2006). As seen in previous source scores those practicing with the easy and medium difficulty levels of the task reported markedly lower levels of negative physiological cues. This is supported by the results of a mixed between-within ANOVA noting a statistically significant difference in physiological source values within difficulty groups between pre and post-test, Wilks-Lambda = .865, F (2,63) = 4.913, p = .01, eta squared = .135. Whether the influence of difficulty on the physiological source is linear in nature requires further research with a larger sample size. The pattern of the sub group practicing with the highest difficulty deviating further from the overall mean than those practicing with the lower difficulty task is also seen in the mastery

and social sources. While this could be accounted for by the relatively small sample size it is worth examining this point with the additional mean values shown in Table 12 (p.105). This raises the question of whether the changes in difficulty settings resulted in a proportional impact on sources. However, the difference between those practicing with the low difficulty and high difficulty is clear. The results support a link between increased difficulty and greater negative physiological source influence. Ultimately increased difficulty resulted in lower self-efficacy levels. When considered in light of the engagement and performance figures; the posited impact of task difficulty on self-efficacy formation is further supported. However, when considering this impact the nature of the relationship between performance and self-efficacy is worthy of consideration.

The Reciprocal Relationship Between Self-Efficacy and Performance

Bandura (1997) theorised that self-efficacy and performance influence each other in a reciprocal manner with a positive outcome leading to increased self-efficacy. This in turn leads to increased engagement, which promotes performance, and the cycle continues (see Figure 7, p.68). However, to date, the majority of studies have focused on between person comparisons of self-efficacy and performance (Williams and Williams 2010). The conclusions drawn from studies that utilise between group comparisons with single data points typically only support a one way effect of self-efficacy enhancing performance. This focuses on the theoretical supposition that an individual's belief about their own capability can enhance their performance. While this is theoretically supported (Schunk and Pajares 2009), it focuses only on the impact of that belief in a once off performance and fails to acknowledge the behaviour altering influence of self-efficacy prior to a performance, such as increased practice attempts. The hypothesised reciprocal relationship (Bandura 1994)

would see a continuous change in self-efficacy levels throughout a practice period based on the individual's interpretation of their own performance. Individuals who perceive a positive outcome from a practice attempt should demonstrate a higher likelihood to engage in further practice attempts due to the self-efficacy enhancing effects of this perceived positive outcome. Conversely perceived negative outcomes would result in lowered self-efficacy and an increased likelihood of disengagement from the practice task. This is supported by the non-significant result of an ANCOVA that examined improvement between difficulty groups (post-test minus pre-test) while controlling for engagement, $F(2,63) = .337$, $p = .715$ which highlights engagement as the main prerequisite of increased performance.

This highlights a tenet of self-efficacy theory that needs to be further empirically explored if the theory is to be further refined. The current study aims to address this gap by presenting data that is the result of a novel research design that facilitates an examination of the nature of the relationship between self-efficacy and performance within a specific task. Firstly the studies examined in Williams and Williams (2010) review failed to account for volition in their design, which in turn does not allow for the examination of the validity of the proposed reciprocal relationship. Self-efficacy is widely accepted to have considerable positive effects on engagement yet the majority of studies only address assessments of performance and accompanying self-efficacy measurements. When repeated attempts at the given tasks were included in the previously mentioned studies (Williams and Williams 2010), they were performed under direct supervision and to completion. This has potentially confounding effects when discussing the relationship between self-efficacy and performance. If a key benefit of increased self-efficacy is greater engagement; then an experiment that removes the option of non-completion from the participant cannot

capture this key element. This is of little use in assessing whether the reciprocal nature of this relationship can be empirically supported. It is worth considering the further complex and dynamic relationship between the sources of self-efficacy and increased engagement. Increased engagement provides increased opportunities for mastery experience, the primary source of self-efficacy. In addition, it is likely to provide more opportunities for social persuasions and vicarious experiences over extended task activities.

The current study examines the proposed reciprocal relationship using data relating to engagement which, when combined with the difficulty manipulation in each round, provides a unique perspective. It allows for the impact of task difficulty on engagement to be examined. Through the hidden varying difficulty settings, used in multiple rounds within this research project, comparisons of difficulty impact can be made. Previous studies have established a clear link between self-efficacy and increased engagement (Bresó *et al.* 2011, van Dinther *et al.* 2011, Joo *et al.* 2013, Parker *et al.* 2014). These studies provided theoretical support for the assertion that task difficulty influenced self-efficacy, and as a result engagement. Comparisons of engagement rates between difficulty groups provide a unique perspective for further examination of this relationship. The divergent nature of engagement rates, in conjunction with the sources of self-efficacy scale data, indicate that task difficulty influenced self-efficacy. The differing rates of disengagement throughout the practice period for each group suggest that the hidden difficulty variations impacted all four sources of self-efficacy. As a result participants' self-efficacy was reduced resulting in an increased likelihood that they would not engage in further practice attempts. This supports the reciprocal nature of the relationship between performance and self-efficacy as repeated manipulation of their perceived

success/failure in the task over time lead to outcomes that are theoretically indicative of lowered self-efficacy. The data from the sources of self-efficacy scale further supports this conclusion and suggests a consistent influence of varying task difficulty across three of the four sources. This relationship is presented graphically, with supporting data, in Figure 24 (p.115).

In summary, the data supports the reciprocal model based on the relationships between engagement rates, performance and self-efficacy source values. The interpretation of the data is further supported by the existing theoretical literature. Ultimately task difficulty manipulations impacted self-efficacy values as participants ascertained their own performance in the practice task. This in turn influenced their likelihood to engage in further practice tasks. The observed variance in engagement and performance is attributed to the influence of task difficulty. Alterations in behaviour aligned with theoretical predictions and demonstrated considerable links to sources of self-efficacy scale data. This raises the question of whether behavioural data variations can be attributed to self-efficacy variances.

Behavioural Manifestations of Self-efficacy Manipulation

Typically research examining self-efficacy tends to focus on its enhancement of certain desirable outcomes such as increased performance or engagement (Bresó *et al.* 2011, Lee 2015, Honicke and Broadbent 2016). Within these studies variance in these desirable outcomes is attributed to variance in self-efficacy values. However, this fails to take account of Bandura's (1997) reciprocal model. This model supports an examination of variance in desirable outcomes (engagement and performance in the case of the current research project) as potential behavioural manifestations of self-efficacy manipulation. While round 5 expressly targets self-efficacy

manipulation by incorporating the sources of self-efficacy scale (Usher and Pajares 2009), it is also argued that earlier rounds provided evidence of self-efficacy manipulation in a less direct manner. Variations in engagement rates and performance are characterised as behavioural manifestations of self-efficacy manipulation. There exists considerable theoretical and empirical evidence that engagement rates and performance are linked to self-efficacy (Williams and Williams 2010, Bandura 2012, Joo *et al.* 2013, Parker *et al.* 2014). However, to date there has been little research examining whether manipulations leading to variance in engagement and performance are reflected in variances in self-efficacy. The earlier rounds of the current study provide empirical evidence of engagement and performance variation similar to that observed in round 5. The previously mentioned theoretical links in addition to round 5 data strongly suggest that variances observed in earlier rounds were indicative of self-efficacy manipulation.

Round 5 data provides additional evidence that variances in engagement rates and performance are linked to manipulations of self-efficacy. Engagement rates demonstrated large positive correlations with mastery (r = .741, n = 66, $p<0.01$), vicarious experience (r = .577, n = 66, $p<0.01$) and social persuasions (r = .693, n = 66, $p<0.01$) (Cohen 1988). In addition a large negative correlation with physiological state source values (r = .-.517, n = 66, $p<0.01$) further supports this point. The alignment between engagement rates and all four sources support the argument that engagement rates are behavioural manifestations of self-efficacy manipulations. The largest observed correlation between mastery and engagement rates is particularly relevant as mastery is typically noted as the primary source of self-efficacy (Aydin and Uzuntiryaki 2009, Usher and Pajares 2009, Williams and Williams 2010). Further support for variances in engagement being linked to variances in self-

efficacy is evident when engagement rates (see Table 8, p.96) and source values (see Table 12, p.105) are compared across difficulty groups.

Overall performance improvement demonstrates a similar relationship with the sources of self-efficacy values from round 5 data (See Table 13, p.113). Performance improvement demonstrated large positive correlations with mastery (r = .708, n = 66, p<0.01), vicarious experience (r = .557, n = 66, p<0.01) and social persuasions (r = .643, n = 66, p<0.01) (Cohen 1988). A similar negative correlation with physiological state was also observed (r = -.508, n = 66, p<0.01). The similar correlation patterns were predictable given the previously highlighted large correlations between engagement and performance in round 5 (r=0.734, n=66, p<0.001). Noting the statistically significant differences between the engagement rates and improvement rates of difficulty groups within round 5, the data strongly indicates that hidden difficulty variations impacted self-efficacy formation. By extension the considerable differences between difficulty groups' engagement and performance figures, and the statistically significant correlations with self-efficacy sources, support the conclusion that variances in engagement and performance figures in this task can be considered behavioural manifestations of self-efficacy manipulations.

CHAPTER SIX: CONCLUSION

Introduction

As outlined in Chapter Three the main aim of the current research project was to examine the relationship between task difficulty, engagement, performance and self-efficacy formation. The impact of task difficulty was examined across multiple rounds with the intention of identifying and quantifying the impact of potentially confounding variables. The results of the first four rounds of the study were examined using Bandura's (1977) self-efficacy theory as a theoretical framework. The final round was linked more directly to self-efficacy theory in its design, incorporating a 'sources of self-efficacy scale' (Usher and Pajares 2009) in order to gain an additional perspective when examining the impact of task difficulty on self-efficacy formation.

The following chapter outlines the primary original contributions of this project towards addressing the previously outlined dearth in the literature surrounding the influence of task difficulty on engagement, performance and self-efficacy formation. The main conclusions are presented using the original research questions as focal points in the earlier stages:

1. What is the relationship between task difficulty and engagement in the given task?
2. What is the associated relationship between task difficulty and performance in the given task?
3. What is the relationship between task difficulty, engagement, performance and self-efficacy formation in the given task?

The latter half of the chapter focuses on recommendations for future research in the area, based on questions arising from this research project.

What is the Relationship Between Task Difficulty and Engagement in the Given Task?

The influence of task difficulty on engagement was noted across all rounds where difficulty was manipulated. In rounds where difficulty was the only manipulated variable the effect sizes calculated using eta squared (round 1: η^2=0.12, round 2 η^2=0.14 and round 5 η^2=0.15) indicate that difficulty had a considerable impact on engagement rates. In all rounds where difficulty was manipulated the group that practiced with the lowest difficulty setting demonstrated the highest engagement rates, while the groups that practiced with the higher difficulty task consistently demonstrated the lowest engagement rates. The results of this study suggest that elevated difficulty levels had a detrimental affect on engagement in the given task. Conversely, lowered difficulty levels demonstrated a positive impact on engagement rates. The results of round 1 demonstrate that this influence is evident outside of a typical school week. In addition, the results of rounds 2 and 5 replicate the findings of round 1 throughout a typical school week.

What is the Associated Relationship Between Task Difficulty and Performance in the Given Task?

A considerable and stable relationship between task difficulty and performance was observed across all rounds where difficulty was manipulated. Groups who practiced with the lower difficulty task demonstrated the greatest performance increases in all rounds. Conversely, groups who practiced with the high difficulty task consistently demonstrated the lowest level of improvement across all rounds where difficulty was

manipulated (See Table 9, p.100). Performance increases demonstrate considerable links to engagement across all rounds. A Pearson's correlation test examining the relationship between practice attempts and improvement, using all five rounds of data, resulted in a large positive correlation r = .525, n = 317, p = .001. This indicates a shared variance of 27.6%. This suggests that the link between difficulty and performance is, at least partly, indirect. While task difficulty exerts a considerable influence on performance in the given task, the performance variances observed demonstrate covariance with engagement. As a result, the observed impact of task difficulty on performance should, in part, be viewed as a product of task difficulty's influence on engagement.

What is the Relationship Between Task Difficulty, Engagement, Performance and Self-Efficacy Formation in the Given task?

In order to fully explore the final research question the relationships between factors are examined in the context of two distinct relationships. Firstly conclusions relating to the relationship between self-efficacy and performance are presented. This is followed by a summation of the impact of difficulty on self-efficacy formation observed within this study.

The relationship between self-efficacy and performance

As discussed in the initial literature review of the current research project, the vast majority of research examining the relationship between self-efficacy and performance focuses on a performance enhancing position. This limits the usefulness of the resultant data when examining the hypothesised bi-directional nature of the relationship between performance and self-efficacy (Williams and Williams 2010). In addition, the performance enhancing studies listed in the earlier review failed to allow for volition in the participants' engagement with the task. Increased self-

efficacy has been linked to increased engagement and decreased burnout (Bresó *et al.* 2011, Zelenak 2015). With the majority of studies limiting their examination to self-efficacy enhancing performance, they fail to provide an opportunity to observe self-determined engagement. Instead a specific number of practice attempts and subsequent performance are enforced. This fails to acknowledge that increased self-efficacy is theoretically linked to increased self-determined engagement which is posited as one way in which self-efficacy can enhance performance (Schunk and Pajares 2009).

The results of this study provide a unique insight into this relationship. The practice period engagement data indicates that those issued with high difficulty versions of the task demonstrated consistently lower engagement rates. Lower engagement rates in round 5 data significantly correlated with the sources of self-efficacy (see Table 13, p.114). This suggests that participants using the higher difficulty task developed lower self-efficacy over time. This can be observed in the disparate engagement rate of each difficulty group. The considerable correlations between self-efficacy source values and performance improvement (see Table 13, p.113) support the assertion that perceived poor performance had a negative impact on self-efficacy. Lowered self-efficacy led to lower engagement rates, which ultimately led to lower overall performance improvement rates. This provides empirical support for the hypothesised bi-directional relationship between self-efficacy and performance; however further research is required in order to more fully understand the intricacies of this relationship.

Ultimately this adds to the considerable body of evidence that self-efficacy is critical to performance. However, the sometimes simplistic manner in which this has been previously examined fails to acknowledge key tenets of self-efficacy theory. While

elevated self-efficacy may have a positive impact on a novel task due to a more positive emotional disposition; the results of this study suggest that the primary source of performance increase is the result of the ongoing reciprocal relationship between performance and self-efficacy discussed previously. The results of this study suggest that engagement rates were the defining factor in terms of final performance. This is important to note as self-efficacy interventions grow in popularity. If increases in performance are the desired result of an intervention then the impact of self-efficacy on behaviour prior to the final performance/assessment should be of greater interest to potential researchers, rather than measures immediately prior to a novel task. In the opening chapter of this thesis self-efficacy was identified as the most broadly studied theory of self-determination. If participants are not afforded the opportunity to direct their own behaviour then the very nature of the theory is undermined.

The relationship between difficulty and self-efficacy formation

The final round of the current study explicitly examined the impact of task difficulty on the four sources of self-efficacy. The results indicate that task difficulty influenced mastery, social persuasions and physiological state. The results did not indicate a distinct influence on vicarious persuasions. Variance observed in mastery source values indicated an effect size of $\eta^2 = .22$ and was found to be large (Cohen 1988). This suggests that increased task difficulty exerted a considerable negative influence on mastery, while conversely decreased task difficulty resulted in increased mastery values. This is supported by the results of a mixed between-within (repeated measures) ANOVA highlighting a statistically significant difference in the change in mastery values within difficulty groups between the pre and post-test (See p.106). Noting that mastery is formed as an individual ascertains their performance

relative to a specific task, it bears the most direct logical and theoretical links to difficulty influence. The data indicates that hidden difficulty variations influenced how individuals perceived their own performance within the task. This suggests that the selection of a difficulty level when designing a task can have considerable impacts on an individual's sense of mastery within that task.

Similar patterns were observed in the social persuasions source data. Variance observed in social persuasions values indicated an effect size of $\eta^2 = .165$ and was found to be large (Cohen 1988). This suggests that increased task difficulty negatively influenced perceived positive social persuasions, while lower task difficulty positively influenced social persuasions values. This is again supported by the results of a mixed between-within (repeated measures) ANOVA highlighting a statistically significant difference in the change in social persuasions values within difficulty groups between the pre and post-test (See p.109). Whether the frequency of social persuasions fluctuated as a result of altered behaviour of the individual attributed to increased difficulty, or whether it was due to an individual perceiving fewer social persuasions remains unclear. Physiological state presents a similar consistent pattern, however, the values are reversed as this source is entirely represented by negative response items in the sources of self-efficacy scale (Usher and Pajares 2009). The findings indicate that increased difficulty resulted in increased negative physiological cues, while lower difficulty resulted in markedly lower negative reports. This is also supported by the results of a mixed between-within (repeated measures) ANOVA highlighting a statistically significant difference in the change in physiological source values within difficulty groups between the pre and post-test (See p.111). This suggests that increased difficulty resulted in negative

emotional responses that are detrimental to the formation of self-efficacy for the given task.

The vicarious experience source did not demonstrate a consistent pattern such as those observed in the other three sources. As highlighted by the developers of the scale used in this study; individuals vary in their interpretation of vicarious experience as a source based on a wide array of personal factors including developmental stage (Usher and Pajares 2009). "This is likely why, with few exceptions, researchers have reported low to modest reliability coefficients among items created to assess vicarious experience" (Usher and Pajares 2009, p.90). Further research examining this source is needed in order to establish whether the inconsistent results are indicative of poor reliability of the scale, or whether the development of this source is more nuanced than the other three previously presented. It is worth noting that a mixed between-within ANOVA highlighted significant differences in changes of this source value within difficulty groups between pre and post-test (See p.107). This raises the possibility that the manner in which difficulty influences the social persuasions source may not be as direct, nor as linear in nature, as the influence of difficulty observed in the other sources.

The distinct influence of task difficulty on mastery, social persuasions and physiological state collectively indicate that task difficulty has a considerable influence on self-efficacy formation. Previously established theoretical links between engagement, performance and self-efficacy support this conclusion (Joo *et al.* 2013, Parker *et al.* 2014, Zelenak 2015). Observed fluctuations in engagement rates and performance increases associated with task difficulty correlate significantly with source values in the final round. This supports the supposition that variations in engagement and performance figures are indicative of manipulated self-efficacy

levels. Ultimately this highlights a unique contribution in the identification of the influence of task difficulty on self-efficacy formation. In addition this study provides data that provides further insight into the reciprocal relationship between performance and self-efficacy as theorised by Bandura (1997), however further research is needed in order to more fully explore this relationship.

Recommendations and Directions for Future Research

This section outlines recommendations arising from the major findings and theoretical implications of this research. As outlined previously, the vicarious source results from the scale used in round 5 suggest a lack of reliability with related validity concerns. This presents an opportunity, and need, for further research. As noted by Liem *et al.* (2008), individuals can base vicarious observances on a great variety of potential models. These can vary from peers, authority figures or parents. It has even been noted that in our ever increasing connection to the digital world models for vicarious experience may be drawn from media sources (Liem *et al.* 2008). The exceptionally diverse nature of potential models is further compounded by the complication of existing relationships with the model. Future studies that could isolate participants from such a diverse range of models. This has the potential to provide insight into this elusive source of self-efficacy. This would require more controlled lab like settings from those used in this study. The findings of such a study would be limited in their generalisability. Nevertheless, it could provide a first step in providing an empirical perspective in what has to date proved to be an elusive, if not unsupported, element of self-efficacy theory.

Linked to a need for further research examining the role of the vicarious source is the need for further investigations of the role of the environment. The results of this

study suggest that the environment in which the task is completed can have a considerable impact on engagement and performance; however this study was limited to one school environment and one round of data outside of this environment. The influence of these environments went beyond the scope of the current study, but the comparison between the first two rounds highlighted considerable variance attributed to the difference between in school and out of school environments. Although the role of environment has been previously linked to self-efficacy formation theoretically and empirically (Caraway *et al.* 2003, Tsai *et al.* 2011, Joo *et al.* 2013), the nuances of this influence remain elusive. Further research examining the role that the environment plays has the potential to also address questions relating to the degree of influence that each source exerts in the formation of self-efficacy. As noted by Schunk and Pajares (2009), the influence of each source can vary due to social and cultural factors. However, empirical investigations regarding these factors are uncommon.

The previous recommendations for areas of further research all called upon future researchers to re-examine various aspects of the current study in order to explore questions relating to self-efficacy theory. This reflects a broader motivation within the field to encourage replication studies. “Confirmation comes from repetition. Any attempt to avoid this statement leads to failure and more probably to destruction” (Tukey 1969, p.84). As highlighted by (Makel *et al.* 2012) the need for replication studies in order to advance aspects of external validity and rigour are vital for the further development of theory.

BIBLIOGRAPHY

Allport, D. A., Antonis, B. and Reynolds, P. (1972) 'On the division of attention: A disproof of the single channel hypothesis', *The Quarterly journal of experimental psychology*, 24(2), 225-235.

Anderson, S. L. and Betz, N. E. (2001) 'Sources of social self-efficacy expectations: Their measurement and relation to career development', *Journal of Vocational Behavior*, 58(1), 98-117.

Annetta, L. A., Minogue, J., Holmes, S. Y. and Cheng, M. T. (2009) 'Investigating the impact of video games on high school students' engagement and learning about genetics', *Computers & Education*, 53(1), 74-85.

Arndt, M. and De Bruin, K. (2011) 'Measurement of attitudes toward bisexual men and women among South African university students: The validation of an instrument', *Journal of homosexuality*, 58(4), 497-520.

Artino, A. R. (2008) 'Motivational beliefs and perceptions of instructional quality: predicting satisfaction with online training*', *Journal of Computer Assisted Learning*, 24(3), 260-270.

Ashford, J. and LeCroy, C. (2000) *Human behavior in the social environment: A multidimensional perspective*, Belmont, CA: Brooks & Cole.

Austin, J. T. and Vancouver, J. B. (1996) 'Goal constructs in psychology: Structure, process, and content', *Psychological bulletin*, 120(3), 338.

Aydin, Y. Ç. and Uzuntiryaki, E. (2009) 'Development and psychometric evaluation of the high school chemistry self-efficacy scale', *Educational and Psychological Measurement* [online], available: doi: doi:10.1177/0013164409332213 [accessed 11 Mar 2016].

Bandura, A. (1977) 'Self-efficacy: Toward a unifying theory of behavioral change', *Psychological Review*, 84(2), 191-215.

Bandura, A. (1978) 'The self system in reciprocal determinism', *American Psychologist*, 33(4), 344.

Bandura, A. (1982) *Social learning theory*, Englewood Cliffs, NJ: Prentice-Hall.

Bandura, A. (1983) 'Temporal dynamics and decomposition of reciprocal determinism: A reply to Phillips and Orton', *Psychological Review*, 90(2), 166-170.

Bandura, A. (1986) *Social foundations of thought and action: A social cognitive theory*, Englewood Cliffs NJ: Prentice Hall.

Bandura, A. (1994) *Self-efficacy*, Hoboken, NJ: John Wiley & Sons, Inc.

Bandura, A. (1995) *Self-efficacy in changing societies*, New York: Cambridge University Press.

Bandura, A. (1997) *Self Efficacy: The Exercise of Control*, New York: Freeman.

Bandura, A. (1999) 'Social cognitive theory of personality' in Pervin, L. A. and Johnson, D. W., eds., *Handbook of personality: Theory and research*, 2 ed., New York: Guilford Publications, 154-196.

Bandura, A. (2006) 'Guide for constructing self-efficacy scales' in Pajares, F. and Urdan, T., eds., *Self-efficacy beliefs of adolescents*, Greenwich, CT: Information Age Publishing, 307-337.

Bandura, A. (2012) 'On the functional properties of perceived self-efficacy revisited', *Journal of Management*, 38(1), 9-44.

Bandura, A., Pastorelli, C., Barbaranelli, C. and Caprara, G. V. (1999) 'Self-efficacy pathways to childhood depression', *Journal of Personality and Social Psychology*, 76(2), 258-267.

Baron, R. S., Vandello, J. A. and Brunsman, B. (1996) 'The forgotten variable in conformity research: Impact of task importance on social influence', *Journal of Personality and Social Psychology*, 71(5), 915-928.

Bates, R. and Khasawneh, S. (2007) 'Self-efficacy and college students' perceptions and use of online learning systems', *Computers in Human Behavior*, 23(1), 175-191.

Battle, E. S. (1965) 'Motivational determinants of academic task persistence', *Journal of Personality and Social Psychology*, 2(2), 209-218.

Beattie, S., Lief, D., Adamoulas, M. and Oliver, E. (2011) 'Investigating the possible negative effects of self-efficacy upon golf putting performance', *Psychology of Sport and Exercise*, 12(4), 434-441.

Betz, N. E. (1978) 'Prevalence, distribution, and correlates of math anxiety in college students', *Journal of Counseling Psychology*, 25(5), 441-450.

Bhaskar, R. (2010) *Reclaiming reality: A critical introduction to contemporary philosophy*, New York: Taylor & Francis.

Blanton, H., Buunk, B. P., Gibbons, F. X. and Kuyper, H. (1999) 'When better-than-others compare upward: Choice of comparison and comparative evaluation as independent predictors of academic performance', *Journal of Personality and Social Psychology*, 76(3), 420-430.

Bong, M. (2001) 'Between-and within-domain relations of academic motivation among middle and high school students: Self-efficacy, task value, and achievement goals', *Journal of educational psychology*, 93(1), 23-31.

Boston, M. D. and Smith, M. S. (2009) 'Transforming secondary mathematics teaching: Increasing the cognitive demands of instructional tasks used in teachers' classrooms', *Journal for Research in Mathematics Education*, 40(2), 119-156.

Bresó, E., Schaufeli, W. B. and Salanova, M. (2011) 'Can a self-efficacy-based intervention decrease burnout, increase engagement, and enhance performance? A quasi-experimental study', *Higher Education*, 61(4), 339-355.

Brewer, M. B. (2000) 'Research design and issues of validity' in Reis, H. and Judd, C., eds., *Handbook of research methods in social and personality psychology*, 2 ed., New York: Cambridge University Press, 3-16.

Britner, S. L. and Pajares, F. (2006) 'Sources of science self-efficacy beliefs of middle school students', *Journal of Research in Science Teaching*, 43(5), 485-499.

Brown, S. L. and Cialdini, R. B. (2015) 'Functional Motives and Functional Consequences of Prosocial Behavior' in Schroeder, D. A. and Graziano, W. G., eds., *The Oxford Handbook of Prosocial Behavior*, New York, NY: Oxford University Press, 346-360.

Buchanan, N. K. and Fox, R. A. (2008) 'Every school a school of choice: school choice in Ireland as viewed through American eyes', *Irish Educational Studies*, 27(3), 267-279.

Bunn, J. (2004) 'Student persistence in a LIS distance education program', *Australian Academic & Research Libraries*, 35(3), 253-269.

Burnette, J. L., O'Boyle, E. H., VanEpps, E. M., Pollack, J. M. and Finkel, E. J. (2013) 'Mind-sets matter: A meta-analytic review of implicit theories and self-regulation', *Psychological bulletin*, 139(3), 655.

Byrne, B. M. (1996) 'Academic self-concept: Its structure, measurement, and relation to academic achievement' in Bracken, B. A., ed., *Handbook of self-concept: Developmental, social, and clinical considerations* Oxford, England: John Wiley & Sons, 287-316.

Campbell, D. T. and Stanley, J. C. (2015) *Experimental and quasi-experimental designs for research*, Ravenio Books.

Caprara, G. V., Fida, R., Vecchione, M., Del Bove, G., Vecchio, G. M., Barbaranelli, C. and Bandura, A. (2008) 'Longitudinal analysis of the role of perceived self-efficacy for self-regulated learning in academic continuance and achievement', *Journal of educational psychology*, 100(3), 525-537.

Caraway, K., Tucker, C. M., Reinke, W. M. and Hall, C. (2003) 'Self-efficacy, goal orientation, and fear of failure as predictors of school engagement in high school students', *Psychology in the Schools*, 40(4), 417-427.

Carver, C. S. and Scheier, M. F. (1982) 'Control theory: A useful conceptual framework for personality–social, clinical, and health psychology', *Psychological bulletin*, 92(1), 111.

Chin, D. and Kameoka, V. A. (2002) 'Psychosocial and contextual predictors of educational and occupational self-efficacy among Hispanic inner-city adolescents', *Hispanic Journal of Behavioral Sciences*, 24(4), 448-464.

Chiou, W.-B. and Wan, C.-S. (2007) 'The dynamic change of self-efficacy in information searching on the Internet: Influence of valence of experience and prior self-efficacy', *The Journal of Psychology*, 141(6), 589-603.

Chiu, C.-M. and Wang, E. T. G. (2008) 'Understanding Web-based learning continuance intention: The role of subjective task value', *Information & Management*, 45(3), 194-201.

Cialdini, R. B. and Trost, M. R. (1998) 'Social influence: Social norms, conformity and compliance' in Gilbert, D. T., Fiske, S. T. and Lindzey, G., eds., *The handbook of social psychology*, 4 ed., New York, NY: McGraw-Hill, 151-192.

Cohen, J. (1988) *Statistical power analysis for the behavioral sciencies*, Hillsdale, NJ: Routledge.

Cohen, L., Manion, L. and Morrison, K. (2000) *Research methods in education*, New York: Routledge.

Cole, J. S., Bergin, D. A. and Whittaker, T. A. (2008) 'Predicting student achievement for low stakes tests with effort and task value', *Contemporary Educational Psychology*, 33(4), 609-624.

Connell, J. P. and Wellborn, J. G. (1991) 'Competence, autonomy, and relatedness: A motivational analysis of self-system processes' in Gunnar, M. R., ed., *Self processes and development*, Hillsdale, NJ: Lawrence Erlbaum Associates, 43-77.

Creswell, J. W. (2013) *Research design: Qualitative, quantitative, and mixed methods approaches*, London: Sage.

Deci, E. L. and Ryan, R. M. (1985) *Intrinsic motivation and self-determination in human behaviour*, New York: Plenum Press.

Department of Education & Science (2005) *Delivering Equality of Opportunity in Schools (DEIS)*, Dublin: Government Publication.

Dipboye, R. L. and Flanagan, M. F. (1979) 'Research settings in industrial and organizational psychology: Are findings in the field more generalizable than in the laboratory?', *American Psychologist*, 34(2), 141-149.

Donaldson, T. S. (1966) *Power of the F-test for nonnormal distributions and unequal error variances*, New York: Rand Corporation.

Eccles-Parsons, J., Adler, T., Futterman, R., Goff, S., Kaczala, C., Meece, J. and Midgley, C. (1983) 'Expectancies, values, and academic behaviors', *Achievement and achievement motivation*, 75-146.

Eccles, J. S. (2005) 'Subjective task value and the Eccles et al. model of achievement-related choices' in Elliot, A. J. and Dweck, C., eds., *Handbook of competence and motivation*, New York: The Guilford Press, 105-121.

Eccles, J. S., Midgley, C. and Adler, T. (1984) 'Grade-related changes in the school environment', *The development of achievement motivation*, 3, 238-331.

Eccles, J. S., O'Neill, S. A. and Wigfield, A. (2005) 'Ability self-perceptions and subjective task values in adolescents and children' in Anderson, K. M. and Lippman, L., eds., *What Do Children Need to Flourish?*, New York: Springer, 237-249.

Eccles, J. S. and Wigfield, A. (1995) 'In the mind of the actor: The structure of adolescents' achievement task values and expectancy-related beliefs', *Personality and Social Psychology Bulletin*, 21(3), 215-225.

Feather, N. T. (1992) 'Values, valences, expectations, and actions', *Journal of Social Issues*, 48(2), 109-124.

Festinger, L., Coren, S. and Rivers, G. (1970) 'The effect of attention on brightness contrast and assimilation', *The American journal of psychology*, 83(2), 189-207.

Fodor, J. A. (1974) 'Special sciences (or: the disunity of science as a working hypothesis)', *Synthese*, 28(2), 97-115.

Fraser, J., Katchabaw, M. and Mercer, R. E. (2014) 'A methodological approach to identifying and quantifying video game difficulty factors', *Entertainment Computing*, 5(4), 441-449.

French, D. P., Olander, E. K., Chisholm, A. and Mc Sharry, J. (2014) 'Which behaviour change techniques are most effective at increasing older adults' self-efficacy and physical activity behaviour? A systematic review', *Annals of Behavioral Medicine*, 48(2), 225-234.

Frese, M. and Zapf, D. (1994) 'Action as the core of work psychology: A German approach' in Triandis, H., Dunette, M. and Hough, L., eds., *Handbook of industrial and organizational psychology*, Paolo Alto, CA: Consulting Psychologists Press, 271-340.

Fröber, K. and Dreisbach, G. (2014) 'The differential influences of positive affect, random reward, and performance-contingent reward on cognitive control', *Cognitive, Affective, & Behavioral Neuroscience*, 14(2), 530-547.

Gainor, K. A. and Lent, R. W. (1998) 'Social cognitive expectations and racial identity attitudes in predicting the math choice intentions of Black college students', *Journal of Counseling Psychology*, 45(4), 403-411.

Geitz, G., Joosten-ten Brinke, D. and Kirschner, P. A. (2016) 'Changing learning behaviour: Self-efficacy and goal orientation in PBL groups in higher education', *International journal of educational research*, 75, 146-158.

Gérard, V. (1998) 'A comprehensive theory of representation for mathematics education', *The Journal of Mathematical Behavior*, 17(2), 167-181.

Gilbert, S. J., Bird, G., Frith, C. D. and Burgess, P. W. (2012) 'Does "task difficulty" explain "task-induced deactivation?"', *Front Psychol*, 3, 125.

Haddad, S. I. and Taleb, R. A. (2016) 'The impact of self-efficacy on performance (An empirical study on business faculty members in Jordanian universities)', *Computers in Human Behavior*, 55, 877-887.

Hamilton, M. (2011) *An Exploratory Study of the Use of Resources by Students With Diverse Levels of Mathematics Self-Efficacy*, unpublished thesis (PhD), Florida State University.

Hampton, N. Z. (1998) 'Sources of Academic Self-Efficacy Scale: An assessment tool for rehabilitation counselors', *Rehabilitation Counseling Bulletin*, 41(4), 260-270.

Hampton, N. Z. and Mason, E. (2003) 'Learning disabilities, gender, sources of efficacy, self-efficacy beliefs, and academic achievement in high school students', *Journal of School Psychology*, 41(2), 101-112.

Harris, J. R. (1995) 'Where is the child's environment? A group socialization theory of development', *Psychological Review*, 102(3), 458-465.

Harter, S. (1996) 'Teacher and classmate influences on scholastic motivation, self-esteem and level of voice in adolescents' in Juvonen, J. and Wentzel, K. R., eds., *Social motivation: Understanding children's school adjustment*, Cambridge, UK: Cambridge University Press, 11-42.

Henson, R. K. (2001) 'Understanding internal consistency reliability estimates: A conceptual primer on coefficient alpha', *Measurement and evaluation in counseling and development*, 34(3), 177-186.

Holden, G., Moncher, M. S. and Schinke, S. P. (1990) 'Self-efficacy of children and adolescents: A meta-analysis', *Psychological Reports*, 66(3), 1044-1046.

Honicke, T. and Broadbent, J. (2016) 'The influence of academic self-efficacy on academic performance: A systematic review', *Educational Research Review*, 17, 63-84.

Hopkins, K. D., Glass, G. V. and Hopkins, B. (1987) *Basic statistics for the behavioral sciences*, London: Prentice-Hall.

Hulleman, C. S., Durik, A. M., Schweigert, S. B. and Harackiewicz, J. M. (2008) 'Task values, achievement goals, and interest: An integrative analysis', *Journal of educational psychology*, 100(2), 398-406.

Hulleman, C. S. and Harackiewicz, J. M. (2009) 'Promoting interest and performance in high school science classes', *Science*, 326(5958), 1410-1412.

Hulleman, C. S., Schrager, S. M., Bodmann, S. M. and Harackiewicz, J. M. (2010) 'A meta-analytic review of achievement goal measures: Different labels for the same constructs or different constructs with similar labels?', *Psychological bulletin*, 136(3), 422-436.

Janis, I. L. (1972) *Victims of groupthink: a psychological study of foreign-policy decisions and fiascoes*, Boston, MA: Houghton-Mifflin.

Jiang, Y., Song, J., Lee, M. and Bong, M. (2014) 'Self-efficacy and achievement goals as motivational links between perceived contexts and achievement', *Educational Psychology*, 34(1), 92-117.

Joo, Y. J., Lim, K. Y. and Kim, J. (2013) 'Locus of control, self-efficacy, and task value as predictors of learning outcome in an online university context', *Computers & Education*, 62, 149-158.

Kanfer, R. and Ackerman, P. L. (1989) 'Motivation and cognitive abilities: An integrative/aptitude-treatment interaction approach to skill acquisition', *Journal of Applied Psychology*, 74(4), 657-668.

Kinderman, T. A., McCollman, T. L. and Gibson, E. J. (1996) 'Peer networks and children's classroom engagements during childhood and adolescents' in Juvonen, J. and Wentzel, K. R., eds., *Social motivation: Understanding children's school adjustment*, Cambridge UK: Cambridge University Press, 279-312.

Klassen, R. M. (2004) 'A Cross-Cultural Investigation of the Efficacy Beliefs of South Asian Immigrant and Anglo Canadian Nonimmigrant Early Adolescents', *Journal of educational psychology*, 96(4), 731-740.

Klein, R. L. (1972) 'Age, sex, and task difficulty as predictors of social conformity', *Journal of Gerontology*, 27(1), 229-236.

Komarraju, M. and Nadler, D. (2013) 'Self-efficacy and academic achievement: Why do implicit beliefs, goals, and effort regulation matter?', *Learning and individual differences*, 25, 67-72.

Kopelman, L. M. (2004) 'Minimal risk as an international ethical standard in research', *The Journal of medicine and philosophy*, 29(3), 351-378.

Kumar, S. and Jagacinski, C. M. (2011) 'Confronting task difficulty in ego involvement: Change in performance goals', *Journal of educational psychology*, 103(3), 664-675.

Lam, S. F., Yim, P. S., Law, J. S. and Cheung, R. W. (2004) 'The effects of competition on achievement motivation in Chinese classrooms', *British Journal of Educational Psychology*, 74(2), 281-296.

Latham, G. P. and Locke, E. A. (1979) 'Goal setting—A motivational technique that works', *Organizational Dynamics*, 8(2), 68-80.

Lee, C. Y. (2015) 'Changes in self-efficacy and task value in online learning', *Distance Education*, 36(1), 59-79.

Lent, R. W., Lopez, F. G. and Bieschke, K. J. (1991) 'Mathematics self-efficacy: Sources and relation to science-based career choice', *Journal of Counseling Psychology*, 38(4), 424-430.

Lent, R. W., Lopez, F. G., Brown, S. D. and Gore Jr, P. A. (1996) 'Latent structure of the sources of mathematics self-efficacy', *Journal of Vocational Behavior*, 49(3), 292-308.

Liaw, S.-S. (2008) 'Investigating students' perceived satisfaction, behavioral intention, and effectiveness of e-learning: A case study of the Blackboard system', *Computers & Education*, 51(2), 864-873.

Liebert, R. M. (1995) *Science and behavior: An introduction to methods of psychological research*, London: Prentice Hall.

Liem, A. D., Lau, S. and Nie, Y. (2008) 'The role of self-efficacy, task value, and achievement goals in predicting learning strategies, task disengagement, peer relationship, and achievement outcome', *Contemporary Educational Psychology*, 33(4), 486-512.

Linnenbrink, E. A. and Pintrich, P. R. (2002) 'Motivation as an Enabler for Academic Success', *School Psychology Review*, 31(3), 313-322.

Lopez, F. G. and Lent, R. W. (1992) 'Sources of mathematics self-efficacy in high school students', *Career Development Quarterly*, 41(1), 3-12.

Lynch, R., Patten, J. V. and Hennessy, J. (2013) 'The impact of task difficulty and performance scores on student engagement and progression', *Educational Research*, 55(3), 291-303.

Maehr, M. L. and Zusho, A. (2009) 'Achievement goal theory' in Wentzel, K. R. and Wigfield, A., eds., *Handbook of motivation at school*, London: Routledge, 77-104.

Magaletta, P. R. and Oliver, J. (1999) 'The hope construct, will, and ways: Their relations with self-efficacy, optimism, and general well-being', *Journal of clinical psychology*, 55(5), 539-551.

Makel, M. C., Plucker, J. A. and Hegarty, B. (2012) 'Replications in psychology research how often do they really occur?', *Perspectives on psychological science*, 7(6), 537-542.

Marsh, H. W. and Craven, R. G. (2006) 'Reciprocal effects of self-concept and performance from a multidimensional perspective: Beyond seductive pleasure and unidimensional perspectives', *Perspectives on psychological science*, 1(2), 133-163.

Martinez, M. (2003) 'High attrition rates in e-learning: Challenges, predictors, and solutions', *The E-Learning Developers' Journal*, 14(1), 38-46.

Matsui, T., Matsui, K. and Ohnishi, R. (1990) 'Mechanisms underlying math self-efficacy learning of college students', *Journal of Vocational Behavior*, 37(2), 225-238.

Meece, J. L. and Miller, S. D. (2001) 'A Longitudinal Analysis of Elementary School Students' Achievement Goals in Literacy Activities', *Contemporary Educational Psychology*, 26(4), 454-480.

Melissinos, C. and O'Rourke, P. (2012) *The art of video games: From Pac-Man to mass effect*, New York: Welcome Books

Messick, S. (1989) 'Validity' in Linn, R. L., ed., *Educational measurement* 3ed., New York: Macmillan, 13-103.

Messick, S. (1994) 'The interplay of evidence and consequences in the validation of performance assessments', *Educational researcher*, 23(2), 13-23.

Messick, S. (1995a) 'Standards of validity and the validity of standards in performance asessment', *Educational measurement: Issues and practice*, 14(4), 5-8.

Messick, S. (1995b) 'Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning', *American Psychologist*, 50(9), 741-750.

Mitchell, M. and Jolley, J. (2012) *Research design explained*, Belmont, CA: Cengage Learning.

Morrow, V. and Richards, M. (1996) 'The Ethics of Social Research with Children: An Overview', *Children & Society*, 10(2), 90-105.

Multon, K. D., Brown, S. D. and Lent, R. W. (1991) 'Relation of self-efficacy beliefs to academic outcomes: A meta-analytic investigation', *Journal of Counseling Psychology*, 38(1), 30-38.

Neuman, W. L. (2005) *Social research methods: Quantitative and qualitative approaches*, Boston, MA: Allyn and Bacon.

Nicholls, J. G. (1984) 'Achievement motivation: Conceptions of ability, subjective experience, task choice, and performance', *Psychological Review*, 91(3), 328-335.

Nordholm, L. A. (1975) 'Effects of group size and stimulus ambiguity on conformity', *The Journal of Social Psychology*, 97(1), 123-130.

Onwuegbuzie, A. J. and Leech, N. L. (2005) 'On Becoming a Pragmatic Researcher: The Importance of Combining Quantitative and Qualitative Research Methodologies', *International Journal of Social Research Methodology*, 8(5), 375-387.

Orvis, K. A., Horn, D. B. and Belanich, J. (2008) 'The roles of task difficulty and prior videogame experience on performance and motivation in instructional videogames', *Computers in Human Behavior*, 24(5), 2415-2433.

Osborne, J. W. and Costello, A. B. (2009) 'Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis', *Pan-Pacific Management Review*, 12(2), 131-146.

Özyürek, R. (2005) 'Informative sources of math-related self-efficacy expectations and their relationship with math-related self-efficacy, interest, and preference', *International Journal of Psychology*, 40(3), 145-156.

Pajares, F. (1997) 'Current directions in self-efficacy research', *Advances in motivation and achievement*, 10(149), 1-49.

Pajares, F. (2003) 'Self-efficacy beliefs, motivation, and achievement in writing: A review of the literature', *Reading &Writing Quarterly*, 19(2), 139-158.

Pajares, F. (2007) 'Empirical properties of a scale to assess writing self-efficacy in school contexts', *Measurement and evaluation in counseling and development*, 39(4), 239.

Pajares, F. and Miller, M. D. (1995) 'Mathematics self-efficacy and mathematics performances: The need for specificity of assessment', *Journal of Counseling Psychology*, 42(2), 190-198.

Pajares, F. and Urdan, T. C. (2006) *Self-efficacy beliefs of adolescents*, Greenwich, CT: Information Age Publishing.

Palmer, D. (2006) 'Sources of self-efficacy in a science methods course for primary teacher education students', *Research in Science Education*, 36(4), 337-353.

Panagos, R. J. and DuBois, D. L. (1999) 'Career self-efficacy development and students with learning disabilities', *Learning Disabilities Research & Practice*, 14(1), 25-34.

Parker, P. D., Marsh, H. W., Ciarrochi, J., Marshall, S. and Abduljabbar, A. S. (2014) 'Juxtaposing math self-efficacy and self-concept as predictors of long-term achievement outcomes', *Educational Psychology*, 34(1), 29-48.

Pervin, L. A. (1992) 'The rational mind and the problem of volition', *Psychological Science*, 3(3), 162-164.

Pintrich, P. R. and Schunk, D. H. (2002) *Motivation in Education: Theory, research and applications*, 2 ed., Upper Saddle River NJ: Merrill/Prentice-Hall.

Pintrich, P. R., Wolters, C. A. and Baxter, G. P. (2000) 'Assessing Metacognition and Self-Regulated Learning' in Schraw, G. and Impara, J., eds., *Issues in the measurement of meta-cognition*, Lincoln, NE: Buros Institute of Mental Measurements, 43-97.

Rosander, M. and Eriksson, O. (2012) 'Conformity on the Internet–The role of task difficulty and gender differences', *Computers in Human Behavior*, 28(5), 1587-1595.

Rosnow, R. L. and Rosenthal, R. (1996) 'Computing contrasts, effect sizes, and counternulls on other people's published data: general procedures for research consumers', *Psychological Methods*, 1(4), 331-339.

Schoon, K. J. and Boone, W. J. (1998) 'Self-efficacy and alternative conceptions of science of preservice elementary teachers', *Science Education*, 82(5), 553-568.

Schunk, D. H. (1987) 'Peer models and children's behavioural change', *Review of Educational Research*, 57, 149-174.

Schunk, D. H. (1989) 'Self-efficacy and achievement behaviors', *Educational Psychology Review*, 1(3), 173-208.

Schunk, D. H. (1995) 'Self-efficacy and education and instruction' in Maddu, J. E., ed., *Self-efficacy, adaptation, and adjustment*, New York: Plenum Press, 281-303.

Schunk, D. H. (2005) 'Self-regulated learning: The educational legacy of Paul R. Pintrich', *Educational Psychologist*, 40(2), 85-94.

Schunk, D. H. and Ertmer, P. A. (2000) 'Self-regulation and academic learning: Self-efficacy enhancing interventions' in Monique, B., Pintrich, P. R. and Moshe,

Z., eds., *Handbook of self-regulation*, San Diego, CA: Academic Press, 783-801.

Schunk, D. H. and Hanson, A. R. (1985) 'Peer models: Influence on children's self-efficacy and achievement', *Journal of educational psychology*, 77(3), 313-322.

Schunk, D. H. and Pajares, F. (2009) 'Self-efficacy theory' in Wentzel, K. and Wigfield, A., eds., *Handbook of motivation at school*, Maryland State, MY: Rotledge, 35-53.

Schunk, D. H. and Zimmerman, B. J. (2012) *Motivation and self-regulated learning: Theory, research, and applications*, Routledge.

Seaborne, A. (1962) 'Group influence on the perception of ambiguous stimuli', *British Journal of Psychology*, 53(3), 287-298.

Shadish, W. R., Cook, T. D. and Campbell, D. T. (2002) *Experimental and quasi-experimental designs for generalized causal inference*, Wadsworth: Cengage learning.

Shaw, C., Brady, L.-M. and Davey, C. (2011) 'Guidelines for Research with Children and Young People', *London: National Children's Bureau Research Centre*.

Sherif, M. (1936) *The psychology of social norms*, New York: Harper.

Sigman, A. (2007) 'Visual voodoo: the biological impact of watching TV', *Biologist*, 54(1), 12-17.

Silvestrini, N. and Gendolla, G. H. (2009) 'The joint effect of mood, task valence, and task difficulty on effort-related cardiovascular response and facial EMG', *International Journal of Psychophysiology*, 73(3), 226-234.

Sitzmann, T. and Ely, K. (2011) 'A meta-analysis of self-regulated learning in work-related training and educational attainment: what we know and where we need to go', *Psychological bulletin*, 137(3), 421.

Sitzmann, T. and Yeo, G. (2013) 'A Meta-Analytic Investigation of the Within-Person Self-Efficacy Domain: Is Self-Efficacy a Product of Past Performance or a Driver of Future Performance?', *Personnel Psychology*, 66(3), 531-568.

Skinner, E. A., Wellborn, J. G. and Connell, J. P. (1990) 'What it takes to do well in school and whether I've got it: A process model of perceived control and children's engagement and achievement in school', *Journal of educational psychology*, 82(1), 22.

Smith, S. M. (2001) 'The Four Sources of Influence on Computer Self-Efficacy', *Delta Pi Epsilon Journal*, 43(1), 27-39.

Stajkovic, A. D. and Luthans, F. (1998) 'Self-efficacy and work-related performance: A meta-analysis', *Psychological bulletin*, 124(2), 240-248.

Stasser, G. and Dietz-Uhler, B. (2008) 'Collective Choice, Judgment, and Problem Solving' in Hogg, M. A. and Tindale, S., eds., *Blackwell Handbook of Social Psychology: Group Processes*, Oxford: Blackwell Publishers Ltd, 31-55.

Steinberg, L., Brown, B. B. and Dornbusch, S. M. (1996) *Beyond the classroom: Why school reform has failed and what parents need to do*, New York: Simon & Schuster.

Stevens, C. K. and Gist, M. E. (1997) 'Effects of self-efficacy and goal-orientation training on negotiation skill maintenance: what are the mechanisms?', *Personnel Psychology*, 50(4), 955-978.

Stevens, T., Olivárez, A. and Hamman, D. (2006) 'The role of cognition, motivation, and emotion in explaining the mathematics achievement gap between Hispanic and White students', *Hispanic Journal of Behavioral Sciences*, 28(2), 161-186.

Thomas, N. and O'kane, C. (2006) 'The ethics of participatory research with children', *Children & Society*, 12(5), 336-348.

Tiku, M. (1971) 'Power function of the F-test under non-normal situations', *Journal of the American Statistical Association*, 66(336), 913-916.

Treiman, D. J. (2014) *Quantitative data analysis: Doing social research to test ideas*, London: John Wiley & Sons.

Tsai, C.-C., Chuang, S.-C., Liang, J.-C. and Tsai, M.-J. (2011) 'Self-efficacy in Internet-based Learning Environments: A Literature Review', *Educational Technology & Society*, 14(4), 222-240.

Tschannen-Moran, M. and Hoy, A. W. (2007) 'The differential antecedents of self-efficacy beliefs of novice and experienced teachers', *Teaching and Teacher Education*, 23(6), 944-956.

Tukey, J. W. (1969) 'Analyzing data: Sanctification or detective work?', *American Psychologist*, 24(2), 83-96.

Urdan, T. (2004) 'Predictors of Academic Self-Handicapping and Achievement: Examining Achievement Goals, Classroom Goal Structures, and Culture', *Journal of educational psychology*, 96(2), 251-264.

Usher, E. L. (2009) 'Sources of middle school students' self-efficacy in mathematics: A qualitative investigation', *American Educational Research Journal*, 46(1), 275-314.

Usher, E. L. and Pajares, F. (2006) 'Sources of academic and self-regulatory efficacy beliefs of entering middle school students', *Contemporary Educational Psychology*, 31(2), 125-141.

Usher, E. L. and Pajares, F. (2008) 'Sources of self-efficacy in school: Critical review of the literature and future directions', *Review of Educational Research*, 78(4), 751-796.

Usher, E. L. and Pajares, F. (2009) 'Sources of self-efficacy in mathematics: A validation study', *Contemporary Educational Psychology*, 34(1), 89-101.

van Daal, T., Donche, V. and De Maeyer, S. (2014) 'The impact of personality, goal orientation and self-efficacy on participation of high school teachers in learning activities in the workplace', *Vocations and learning*, 7(1), 21-40.

van Dinther, M., Dochy, F. and Segers, M. (2011) 'Factors affecting students' self-efficacy in higher education', *Educational Research Review*, 6(2), 95-108.

Wang, A. Y. and Newlin, M. H. (2002) 'Predictors of web-student performance: The role of self-efficacy and reasons for taking an on-line class', *Computers in Human Behavior*, 18(2), 151-163.

Wang, C.-H., Shannon, D. M. and Ross, M. E. (2013) 'Students' characteristics, self-regulated learning, technology self-efficacy, and course outcomes in online learning', *Distance Education*, 34(3), 302-323.

Weber, M., Shils, E., Finch, H. A. and Antonio, R. J. (2011) *Methodology of Social Sciences*, Piscataway, NJ: Transaction Publishers.

White, R. W. (1959) 'Motivation reconsidered: The concept of competence', *Psychological Review*, 66(5), 297-333.

Wigfield, A. and Eccles, J. S. (1992) 'The development of achievement task values: A theoretical analysis', *Developmental Review*, 12(3), 265-310.

Williams, T. and Williams, K. (2010) 'Self-efficacy and performance in mathematics: Reciprocal determinism in 33 nations', *Journal of educational psychology*, 102(2), 453-464.

Wolters, C. A., Yu, S. L. and Pintrich, P. R. (1996) 'The relation between goal orientation and students' motivational beliefs and self-regulated learning', *Learning and individual differences*, 8(3), 211-238.

Wood, R. and Bandura, A. (1989) 'Social cognitive theory of organizational management', *Academy of management review*, 14(3), 361-384.

Wood, W. (2000) 'Attitude change: Persuasion and social influence', *Annual review of psychology*, 51(1), 539-570.

Woolfol, A. E. and Hoy, H. K. (1990) 'Prospective teachers' sense of efficacy and beliefs about control', *Journal of educational psychology*, 82, 81-91.

Zelenak, M. S. (2015) 'Measuring the Sources of Self-Efficacy Among Secondary School Music Students', *Journal of Research in Music Education*, 62(4), 389-404.

Zimmerman, B. J. (1990) 'Self-regulated learning and academic achievement: An overview', *Educational Psychologist*, 25(1), 3-17.

Zimmerman, B. J. (2000) 'Self-Efficacy: An Essential Motive to Learn', *Contemporary Educational Psychology*, 25(1), 82-91.

Zimmerman, B. J. and Bandura, A. (1994) 'Impact of Self-Regulatory Influences on Writing Course Attainment', *American Educational Research Journal*, 31(4), 845-862.

Zimmerman, B. J., Bandura, A. and Martinez-Pons, M. (1992) 'Self-Motivation for Academic Attainment: The Role of Self-Efficacy Beliefs and Personal Goal Setting', *American Educational Research Journal*, 29(3), 663-676.

Zimmerman, B. J. and Martinez-Pons, M. (1990) 'Student differences in self-regulated learning: Relating grade, sex, and giftedness to self-efficacy and strategy use', *Journal of educational psychology*, 82(1), 51-59.

List of Appendices:

Appendix A: Research Information and Consent Documents

Parental/Subject Information Sheet.

Project: The Pacman Study

Aim of Research: To assess the impact of task difficulty on participant performance

This research aims to assess the impact of the perceived challenge of a task on student motivation, engagement, and ultimately performance in that task. For the purpose of this study participants will be required to complete a computerised maze navigation task in the form of the enduring Pacman computer game.

The objective of the study is to determine whether or not the perceived difficulty of the Pacman game has an impact on participant motivation, engagement and ultimately performance over a one week period practicing with the game. As part of the study participants will be asked to play the Pacman game under supervision and to register their highest score in the game. Subsequently, participants will be given a version of the game to practice with over the course of a school week, as they see fit. Participants will be invited to replay the game under supervision at the end of the school week to try and beat their previous top score. In order to reduce impact and to comply with industry guidelines the game should not be played for more than one hour a day.

Please find attached a consent form requesting your participation in this research project. Please read and indicate your willingness to grant consent for participation in the study by completing the attached form. If you have any questions or problems with the supplied software please feel free to contact me at any stage.

Participants are voluntarily partaking in this study and at any time may choose to withdraw from the study. ***At no stage will a participant's name ever be used or published in any document.***

Jason Power

Jason Power

Postgraduate Researcher

University of Limerick

Email: jason.power@ul.ie

Phone: 0877608640

University *of* Limerick

OLLSCOIL LUIMNIGH

Consent Form for

The Pacman Study

Dear Sir/Madame,

As part of an education research project, I am conducting a study which aims to assess the impact of the perceived challenge of a task on student motivation, engagement, and ultimately performance in that task. In this study the task is the Pacman computer game.

Before signing this form, you must be provided with details of this study in which you may give consent to participate in. Participation is voluntary and one can withdraw from the study at any stage without giving a reason. The questionnaire is completely anonymous and the department guarantees that if one decides to participate, all of resulting data will be kept confidential. No unauthorised individuals will have access to it and any forms concerning participant data will be destroyed after the study has been completed. Your cooperation in this research would be greatly appreciated.

Yours sincerely,

Jason Power__

Jason Power

University of Limerick

I, _____, have read the subject information sheet regarding the outlined study in which I am about to give consent to participate in. I am aware of all of the procedures involved, and, I understand that participants can withdraw from the study at any time.

Date: _____ Signed: _____

Parents/carers signature: _____

If you have concerns about this study and wish to contact someone independent, you may contact:

Chair,

Faculty of Science & Engineering Research Ethics Committee,

University of Limerick,

(061) 202802

Appendix B: Sources of Self-efficacy Scale

Adapted Sources of Self-efficacy Scale

Item	Full Sample	
1	I make excellent scores on Pacman(ME-1)	Mastery
2	I am successful on Pacman (ME-3)	
3	Even when I practice very hard, I do poorly in Pacman (ME-6)	
4	I got good scores in Pacman on my practice games (ME-8)	
5	I do well on Pacman (ME-9)	
6	I do well on even the higher levels of Pacman (ME-12)	
7	Seeing others do well in Pacman pushes me to do better (VA-4)	Vicarious Experience
8	When I see how others win a game, I can picture myself winning the game in the same way (VA-6)	
9	Seeing others do better than me in Pacman pushes me to do better (VP-1)	
10	When I see how another student wins a game of Pacman, I can see myself winning in the same way (VP-9)	
11	I imagine myself completing Pacman successfully (VS-4)	
12	I compete with myself in Pacman (VS-5)	
13	Others have told that I am good at playing Pacman (P-4)	Social Persuasions
14	People have told me that I am good at Pacman (P-5)	
15	Adults in my family have told me that I am good at Pacman (P-7)	
16	I have been praised for my ability in Pacman(P-13)	
17	Other students have told me that I'm good at playing Pacman (P-14)	
18	My classmates think I'm good at Pacman (P-16)	
19	Playing Pacman makes me feel stressed and nervous (PH-2)	Physiological State
20	Practicing Pacman takes all of my energy (PH-3)	
21	I start to feel stressed-out as soon as I begin playing Pacman (PH-5)	
22	My mind goes blank and I am unable to think clearly when playing Pacman (PH-7)	
23	I get depressed when I think about playing Pacman(PH-9)	
24	My whole body becomes tense when I have to play Pacman (PH-12)	

*Codes listed after each item correspond with the original items developed and validated by Usher and Pajares (2009). Adaptations included replacing terms related to mathematics based tasks with references to the computerised maze navigation task presented as 'Pacman'. In addition references to 'math teachers' were replaced with 'others' in item 13 due to teachers having no involvement, or knowledge of performance, in the given task.

Original Sources of Self-efficacy Scale items

Item	Full Sample	
1	I make excellent grades on math tests (ME-1)	Mastery Experience
2	I have always been successful with math (ME-3)	
3	Even when I study very hard, I do poorly in math (ME-6)	
4	I got good grades in math on my last report card (ME-8)	
5	I do well on math assignments (ME-9)	
6	I do well on even the most difficult math assignments (ME-12)	
7	Seeing adults do well in math pushes me to do better (VA-4)	Vicarious Experience
8	When I see how my math teacher solves a problem, I can picture myself solving the problem in the same way (VA-6)	
9	Seeing kids do better than me in math pushes me to do better (VP-1)	
10	When I see how another student solves a math problem, I can see myself solving the problem in the same way (VP-9)	
11	I imagine myself working through challenging math problems successfully (VS-4)	
12	I compete with myself in math (VS-5)	
13	My math teachers have told that I am good at learning math (P-4)	Social Persuasions
14	People have told me that I have a talent for math (P-5)	
15	Adults in my family have told me what a good math student I am (P-7)	
16	I have been praised for my ability in math (P-13)	
17	Other students have told me that I'm good at learning math (P-14)	
18	My classmates like to work with me in math because they think I'm good at it (P-16)	
19	Just being in math class makes feel stressed and nervous (PH-2)	Physiological State
20	Doing math work takes all of my energy (PH-3)	
21	I start to feel stressed-out as soon as I begin my math work (PH-5)	
22	My mind goes blank and I am unable to think clearly when doing math work (PH-7)	
23	I get depressed when I think about learning math (PH-9)	
24	My whole body becomes tense when I have to do math (PH-12)	

*Original items of the 'Sources of Self-efficacy Scale' (Usher and Pajares 2009)

Appendix C: Data Disc