The numbing effect of experience

by

Frank Giordano

B.S., Manhattan College, 2014

A THESIS

submitted in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

Department of Psychological Sciences
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2019

Approved by:

Major Professor
Jin Lee, Ph. D

Copyright

Abstract

Job experience is a pervasive metric used in human resource functioning; however, its predictive validity might not be as intuitive as it seems especially regarding safety outcomes. While research suggests a positive relationship between experience and performance (McDaniel, Schmidt, & Hunter, 1988; Sturman, 2003), there are instances when experience may have a null or negative relationship with performance (Woltz et al., 2000). Specifically, this occurs when the perceived similarity between prior experience and a new task is actually discrepant. Also, rigidities in problem-solving can form from job experience, leading to impaired performance (Dokko, Wilk, & Rothbard, 2009; Bilalić, McLeod, & Gobet, 2008).

The relationship between job experience and safety outcomes is more intricate. Self-appraisals of safety performance can be discrepant, which is problematic when tasks include safety behaviors that individuals assume they can adequately enact. The means of informing self-efficacy can also be inequitable, such that positive safety behavior feedback (e.g., no injury or accident) can go ungiven or unnoticed, while performance feedback is often a focus of organizations (Gun, 1993). This might contribute to the false perception of similarity in discrete job tasks and safety behaviors (i.e., task performance vs. safety performance). Individuals can be exposed to work experiences which do not provide adequate opportunities to discern work system components, inherent hazards, and risks. This study aims to observe the effect prior experience has on performance of a task, the execution of safety behaviors, and participants' appraisals of their self-efficacy regarding both the task and their safety behaviors.

Table of Contents

List of Figures

List of Tables

Acknowledgements

This would not have been possible without the constant support from my colleagues, friends, and family. I would like to thank my advisor, Dr. Jin Lee, for his guidance and patience. I would also like to thank my committee members, Dr. Chris Lake and Dr. Patrick Knight, for their valuable input. Finally, I would like to thank my cohort (+1), Stacy, Kelsey, Nathan, and Cassie. Your friendship and encouragement were paramount in maintaining my sanity throughout all of the graduate school experience, especially during this thesis.

Chapter 1 - Introduction

Job experience is a pervasive metric often used in human resource functioning; however, its predictive validity might not be as intuitive as it may seem especially regarding safety outcomes. While research suggest a positive relationship between experience and performance (McDaniel, Schmidt, & Hunter, 1988; Sturman, 2003), there are instances when experience may not have an effect or even share a negative relationship with performance (Singley & Anderson, 1989; Woltz et al., 2000). Specifically, this occurs when there is a discrepancy between the perceived similarity between experience and a new task (Gick & Holyoak, 1987). Also, rigidities in problem-solving and a loss of dynamic thought can form from job experience that can lead to impaired performance (Dokko, Wilk, & Rothbard, 2009; Bilalić, McLeod, & Gobet, 2008a; Bilalić, McLeod, & Gobet, 2008b).

When it comes to safety performance, the relationship between job experience and safety-related behaviors/outcomes is even more intricate. Experience can yield the requisite cognition and social factors necessary to inform self-efficacy, and consequently other forms of self-evaluation (Shelton, 1990; Sherer et al. 1982; Eden & Aviram, 1993). These self-appraisals can also be discrepant while influencing one another, which is problematic when tasks include safety behaviors that individuals assume they can adequately enact. The means of informing self-efficacy can also be inequitable, such that positive safety behavior feedback (e.g., no injury or accident) can often go ungiven and unnoticed, while performance feedback is often given more noticeably and consistently. This might contribute to the false perception of similarity in discrete job tasks and safety behaviors (i.e., task performance vs. safety performance) and thus inaccuracy of self-appraisals regarding task and safety self-efficacy (i.e., “I am good at my work.” vs. “I am a safe worker.”). Individuals can be exposed to work experience which does not provide adequate opportunities to discern various work system components, inherent hazards, and risks. They may conflate their job self-efficacy with their abilities to enact safety behaviors, potentially numbing them to occupational hazards and risks. To this end, this study aims to observe the effect prior task experience has on performance of a task, the execution of safety behaviors while performing the task, and participants' appraisals of their self-efficacy regarding both the task and their safety behaviors.

The Demand for Safety Research

In 2016 alone, 2.9 million nonfatal work injuries and 5,190 fatal work injuries were sustained by American civilian workers (BLS, 2016). This marked the third consecutive annual increase in fatal work injuries, and was the first breach of 5,000 fatalities due to work injury since 2008. These fatal injuries included: transportation incidents; falls, slips, and trips; contact with objects and equipment; exposure to harmful substances or environments; and fires and explosions, which are arguably avoidable injuries. Furthermore, about 20% of total fatal workplace injuries are incurred by foreign-born workers, with 37% of these workers being born in Mexico. Asian and Black workers suffered an increase in injury frequency (40% and 19 % respectively) from 2015 (BLS, 2016). Workplace safety and health behaviors need constant attention and maintenance in order to provide benefit to workers. This is especially true for vulnerable populations, as immigrants and non-whites tend to be employed contingently at a higher rate, which can leave them with fewer psychological and tangible resources such as job control, pay, and job training (Benach & Muntaner; 2007; Lewchuk, de Wolff, King, & Polanyi, 2003; Rousseau & Libuser, 1997).

In order to promote and maintain workplace safety behaviors, particular strategies such as safety training and intervention programs can be employed. These interventions include ergonomic improvements (Yeow & Shen, 2003), safety behavior and knowledge training (Burke, Sarpy, Smith-Crowe, Chan-Serfin, Salvaor, & Islam, 2006), salience training (DeJoy, Smith, & Dyal, 2017; Ajzen, 2002), and improvements in safety climate (Mullen & Kelloway, 2009; Sokas et al., 2009). The caveat to these methods is that they cost resources, particularly money and time. For instance, the cost for the updated version of the Occupation Safety and Health Administration (OSHA)'s basic 10-hour construction training course is $89 per person (OSHA, 2018), which can be required for employment. These OSHA packages can increase in price to $305 depending on the nature of the training and the amount of credit hours necessary to achieve the specific certification. Due to these costs, it has been documented that some organizations will avoid providing training when possible in order to save on organizational costs, especially in situations where turnover is expected, as is the case with providing safety training to temporary workers (Rousseau & Libuser, 1997). Significant organizational support and engagement, particularly from superiors, is also necessary for these strategies to provide results. These strategies will often employ the manipulation of safety climate, which is conceptualized as the shared perception and value of safety in an organization that is expressed through policies, procedures, and practices of supervisors and peers (Zohar, 1980). Without a concerted effort from management to creating positive cultural shifts, these interventions will not enact any change.

Instead of costly safety training or cultural improvements, organizations may rely on selection procedures to provide them with capable and safe workers, as opposed to cultivating this from within. The use of previous work experience in the same or similar working environment as a proxy comes with the hope that workers may have developed safety knowledge and skills elsewhere. Research suggests that this can come with caveats, such as rigidities, lack of adaptability, and decreased initial performance (Dokko, Wilk, & Rothbard, 2009; Bidwell, 2011). Ignorance to management's responsibility to ensure workplace safety coupled with the enforcement of beliefs that safety issues exist primarily at the individual level can lead to instances in which blame is shifted to workers for negative safety incidents, although they were not adequately educated, trained, empowered, equipped, or protected to address the safety issues.

The present study attempts to answer whether experience and the associated self-efficacy beliefs can yield similar increases in performance in both task and safety domains as well as examine any systematic alterations of appraisals of task self-efficacy and safety self-efficacy due to experience and performance. To this end, this study aims to examine the relationship between experience and appraisals of self-efficacy, as well as the relationship between self-efficacy appraisals and task performance and the performance of ancillary safety behaviors.

Operationalization of Experience and Implications for Performance

Previous experience is often used as a ubiquitous metric in terms of human resource management, specifically training (Ford, Quinones, Sego, & Sorra, 1992; Blume, Ford, Baldwin, & Huang, 2010), career development (Campion, Cheraskin, & Stevens, 1994), and selection (Ash & Levince, 1985; Rowe, 2018). Early research suggests that the construct of experience is capturing job-relevant knowledge accrued over time (Fiedler, 1970). Quinones, Ford, and Teachout (1995) argue that this may be the case; however, a distinction between different dimensions of knowledge must be made. According to the authors, experience would be a greater indicator of procedural knowledge gained than other types of knowledge. They also note that individual differences can moderate the quality of experience received from a particular event, such that contextual factors surrounding experience should be examined.

Early research mirrors this caution when using experience to predict performance. Fiedler (1970) suggests that the efficacy of experience as a predictor of performance should be brought into question, prompting further research on the construct and the contextual factors in which experience can be gained. In a meta-analysis, Hunter and Hunter (1984) find a modest correlation ($r = .18$) between hiring managers' perceptions of relevant job experience and job performance, which aligns with Fielder's concerns. Schmidt, Hunter, and Outerbridge (1986) found that job experience had a direct effect on job performance when assessed by a job sample, when the relevant job experience is greater than two years, and when the job being considered is of intermediate complexity. After these considerations were made, McDaniel, Schmidt, and Hunter (1988) found a mean corrected correlation of .32 between job experience and performance in their meta-analysis that included a number of varied occupations. While this result serves as evidence for the use of experience as a predictor of job performance, conclusions about the usefulness of experience as a predictor of performance remain mixed. Research has shown the existence of null effects and even negative effects on performance (Castilla, 2005; Medoff & Abraham, 1980), indicating that the relationship between experience and performance should be closely examined on a contextual basis.

It is posited that these mixed results come from the varying operationalizations of experience and the contextual factors surrounding experience (Dokko, Wilk, & Rothbard, 2009). Quinones et al. (1995) mention that the operationalization of experience and contextual factors of experience must be considered for the fair comparison of experience across individuals. Specifically, common methods of quantifying experience include age, company tenure, years spent in a particular industry (industry tenure), and amount of times a discrete task in completed. Measures that use the same unit (e.g. years) are still incomparable due to contextual factors, making tenure, age, and years in an industry not equivalent. Even heterogeneous measures of experience might be confounded with numerous individual and organizational factors that can attenuate the quality of experience such as attitude, motivation, and leadership. While these different measures of experience should not be directly compared, multiple unique operationalizations that have been used in research and are considered equal when discussing experience as a predictor. For instance, age, company, and industry tenure have been used as a proxy for experience in order to estimate potential for performance and extra role behaviors, such as occupational citizenship behaviors (Ng & Feldman, 2008; Ng & Feldman, 2009, Waldman & Avolio, 1986). Researchers have also developed instruments in order to assess how much domain experience an individual might have in the cases of leadership roles and global work experience in order to predict efficacy as a leader (DeRue & Wellman, 2009; Dragoni, Vankatwyk, & Tesluk, 2011). This wide variety of legitimate operationalizations makes comparison of experience across studies difficult to achieve.

In their meta-analysis, Quinones at al. (1995) established support for the multidimensionality of experience by compiling past research and organizing the previous operationalizations of experience using two dimensions, measurement mode and level of specificity (Figure 1). Mode indicates the measurement scheme being used to evaluate time, either by discrete number of tasks performed, the time spent performing job tasks, and type of the experience received. Level of specificity denotes the explicit point (task, job, or organization) in which the experience is measured in the organization. Quinones et al. use a 3 x 3 explanation to create a conceptual framework to address the multiple operationalizations and compare their efficacy. For example, tenure within a company would be considered a time measurement mode at the organization level of specificity.

The same meta-analysis displays that all posited forms of experience have a positive relationship with performance; however, some modes are used more frequently than others, and have varying levels of predictive ability (Quinones et al., 1995). Concerning measurement mode, 79.5% of studies used a time-based measurement mode of experience, 11.4% used the amount of complete tasks mode, and 9.1% categorize experience into types, including supervisory, new recruit, and instructors. For level of specificity, 13.6% of studies measured experience at the task level, 68.2% measured experience at the job level, and 18.2% at the organizational level of specificity. Quinones et al. found that performance has a positive relationship with both soft measurements (self-report data, supervisor rating, etc.) and hard measurements (work samples) of experience; however, these relationships varied across mode and specificity. The strongest relationship between performance and measurement mode of experience was amount of times a task was completed (M_ρ = .43; SD_ρ = .17; $SE_{M\rho}$ = .03), followed by time doing said task (M_ρ = .27; SD_ρ = .11; $SE_{M\rho}$ = .01), and then type of task (M_ρ = .21; SD_ρ = .00; $SE_{M\rho}$ = .05). The strongest relationship between performance and level of specificity of experience was the task level (M_ρ = .41; SD_ρ = .17; $SE_{M\rho}$ = .03), followed by job level (M_ρ = .27; SD_ρ = .12; $SE_{M\rho}$ = .01), and then organization level (M_ρ = .16; SD_ρ = .20; $SE_{M\rho}$ = .04). While tenure measured at the job level is the most common form of measuring experience; it is not the most valid in terms of predicting performance. The authors mention that greater level of specificity to discrete job tasks not only has the strongest predictive ability among the operationalizations of experience, but also matches conceptually with what most research attempts to achieve when discussing job experience.

Research supports the claim that contextual factors matter greatly, and can attenuate or improve the effect job experience has on performance. Sturman (2003) suggests that the relationship between experience and job performance is not equally generalizable across all jobs. Different contexts and complexities change the existing relationship, such that the predictive ability of experience is strongest in jobs with high complexity. For example, work experience can significantly predict performance when considering job tasks that require complex understanding of arithmetic or logical problem-solving (Kolz, McFarland, & Silverman, 1998). The general consensus is that while experience with a task eventually leads to improvement, context and complexity moderate the relationship that exists between experience and job performance, and each case should be observed independently.

H1a: Experience, determined by trial, will be positively related to task performance.

Implications of Experience for Safety Performance

The primary focus of the present study is that the contextual and cognitive mechanisms that are accrued along with experience may not be as beneficial for promoting workplace safety performance as they are for the promotion of job performance. The aforementioned pitfalls that accompany the use of job experience as a predictor of desirable work outcomes can be prevalent regarding safety behaviors and outcomes. While employees might be gaining job-relevant knowledge over time, they may also adopt norms, form cognitive scripts, and ultimately create habits that may not be conducive to all job tasks, especially safety behaviors. To better understand these mechanisms and how task performance-related constructs may inadvertently be conflated with safety behaviors and safety self-efficacy, the relationships between experience, task performance, and self-efficacy need to be investigated.

Expectations regarding behaviors and institutional norms regarding how work is to be completed can be acquired through experience (Beyer & Hannah, 2002; McCall, 1990). Industry-level norms can become habituated while working in an industry (Chatman & Jehn, 1994). Experience can impart more than knowledge to job incumbents, resulting in work habits and lasting attitudes that are not congruent with their given work context. Experience can also lead to cognitive scripts, which are cognitive patterns that organize and encode information and guide subsequent behaviors. Work often includes completing the same job task repeatedly in the same environment over an extended period of time, meeting the prerequisites for enforcing scripts (Gioia & Poole, 1984; Markus & Zajonc, 1985; Walsh, 1995).

Misunderstanding of similar tasks can inadvertently result in sub-optimal transfer of experience, or the learning and misguided use of previously acquired knowledge to different domains of performance (Singley & Anderson, 1989; Woltz et al., 2000). This can occur when situations are appraised as similar, but underlying cultural, industrial, or structural differences exist (Novick, 1990; Woltz at al., 2000). When encountering a new task, previous successful strategies from experiences in similar but distinct domains are often used in a process known as analogical transfer. Analogical transfer can unwittingly result in negative transfer when surface features of the problem appear similar yet the structural features are different, and when an individual is a novice (Novick, 1990).

Research also supports that once task-relevant knowledge and skill are controlled for, experience can have a negative effect on performance, and that lower levels of adaptability can attenuate this effect (Dokko et al., 2009). Archival data attained from an insurance agency displayed that agents performed worse when they had greater levels of experience after relevant job knowledge was controlled. Agents with greater rigidity (less adaptability) were more susceptible to this effect. Rigidity in problem solving, known as the Einstellung effect, can also attribute to a lack of performance and critical thought (Bilalić, McLeod, & Gobet, 2008a; Bilalić, McLeod, & Gobet, 2008b). Chess players were tasked with solving problems that had both a suboptimal familiar solution and a more optimal unfamiliar solution, in which players defaulted to the familiar option. The Einstellung effect was seen even in experts in the presence of the familiar solution; however, this effect was weakened by greater levels of expertise (Bilalić, McLeod, & Gobet, 2008b). This is also supported by research comparing internal promotion and external hiring, suggesting that external hires will have worse initial performance when placed into a new organization (Bidwell, 2011).

These examples display some of the potential contextual factors that may hamper the predictive validity of experience on constructs besides performance, particularly in situations that are not clearly defined. Workplace safety performance is worth of investigation as one of these unclear situations for following reasons. While task and safety behaviors may be enacted in tandem, task experience and safety experience can be different. For example, effective behavior may not always align with safe behavior. Also, efforts necessary for safer performance are not always compatible with task performance. In the case of using a painter using a ladder, an individual may perform well while not wearing proper safety equipment, hanging tools and paint from the ladder for easy access, and standing on the top rung; however, all these actions are considered unsafe. Paying extra attention while on the ladder, moving slowly and cautiously, and having another worker stabilize the ladder situation can slow down and compromise one’s and other’s task performance.

The assumption that workers with more experience are more likely to perform better in terms of both work and safety can be erroneous, considering potential irrelevance and incompatibility between performance, safety. and prior job experience. It is worth noting that many organizations and managers often place less significance on matters of safety in relation to organizational success (Kartam, Flood, & Koushki, 2000). This is more likely to happen in the face of industrial pressures such as deadlines closing in and projects going over budget (Gun, 1993). This also results in reliance of contingent workers and subcontractors, which are often associated with decreased safety conditions and lack of safety training (Quinlan & Soukas, 2009; Gun, 1993). When safety training and interventions are given, they are often seen as cursory, poorly outsourced, and lacking in explicit support from leadership (Hansen, 1993). This greater emphasis on overall organizational performance (e.g., meeting deadline, number of products assembled, remaining under budget) than on safety performance (e.g., compliance with safety rules, prevention of accident/injury) may also manifest at the individual level. Considering this common trend of disproportionate prioritization of productivity in business, it can be inferred that task performers' vigilance to safety features is likely to be compromised across recurring phases of task and safety experience.

H1b: Experience, determined by trial, will have a stronger positive relationship with task performance than safety performance.

Implications of Experience and Self-Efficacy for Task Performance

Self-efficacy is a mechanism in which people generate beliefs in their abilities and expectations about their performance on a specific task, and is typically assessed from prior experience (Bandura, 1977; Bandura, 1982). These beliefs can vary in their level of difficulty in which the task can be successfully performed (magnitude), the certainty regarding successful performances (strength), and the generalizability of the magnitude and strength beliefs across different tasks (generality) (Bandura, 1997). Higher levels of self-efficacy have been associated with a host of positive work-related outcomes, such as performance, civility, and motivation (Zimmerman, 1990; Bandura, 1997; Rhee, Hur, & Kim 2017). For instance, self-efficacy has been used to predict job satisfaction (Judge & Bono, 2001), reductions in job stress (Rennesund & Saksvik, 2010), and work engagement (Consiglio, Borgogni, Di Tecco, & Schaufeli, 2016). Self-efficacy has been shown to predict job performance, especially when the complexity of the job is low (Judge & Bono, 2001; Judge, Jackson Shaw, Scott & Rich, 2007). Research conducted by Pajares and Kranzler (1995) displays a direct effect of self-efficacy on mathematical ability that had incremental predictive validity above and beyond general mental ability. This being said, the cultivation of self-efficacy should be a goal organizations strive to achieve in order to provide their employees with the previous host of positive outcomes.

Interventions are often conducted to increase self-efficacy in a workforce to induce these positive outcomes and foster career development (Gist & Mitchell, 1992; Hackett & Betz, 1981; Tia, 2006). In order to increase self-efficacy, appropriate information on performance must be delivered to the individual through multiple internal and external sources. Bandura (1986) provides a framework known as social cognitive theory, which states that self-efficacy is generated from both cognitive and social sources, and that this self-perception of ability can trigger behavioral change. These sources include mastery experiences, vicarious experiences, and social persuasion. Mastery experiences are instances in which individuals engage in a task, appraise their own performance, and make judgments on their ability. Beliefs of self-efficacy will then be raised or lowered depending on whether the individual receives the desired outcome. Vicarious experiences involve social comparisons and peer modeling, in which an individual will learn competence from others' actions and model their own behavior accordingly. Social persuasion comes in the form of feedback from others that becomes internalized, altering the individual's perception of their own self-efficacy. Despite their varying strengths, combinations of these sources are considered by individuals while making self-efficacy appraisals. Schunk (1981, 1987) has shown that while mastery experience is the strongest method of altering self-efficacy beliefs, and is often the primary method of doing so, with experience with a task service as practice and a method to gauge ability.

H2a: Experience, determined by set of trials (section), will be positively related to task self-efficacy.

H3a: Greater levels of task performance will be related to greater levels of both task and safety self-efficacy, with a greater relationship with task self-efficacy

Implications of Experience and Self-Efficacy for Safety Performance

The mastery, vicarious, and social experiences requisite for the increase of task performance may not be the same as those for increasing ancillary performance, such as safety. Research conducted by Katz-Navon, Naveh, and Stern (2007) proposed a safety self-efficacy construct that can address this discrepancy. They surveyed nurses on enactive mastery experiences; managers as safety role models; verbal persuasion; and an additional contributor unique to safety self-efficacy, safety priority. Results showed that managers serving as safety role models; persuasion via the distribution of safety information; and safety priority contributed to safety self-efficacy. While this is the case, performance was operationalized as patient safety, using a nine-item questionnaire. Patient safety is a primary goal of nursing, and avoiding unnecessary harm to patients is a core portion of their job. Other instances of safety at work can include one's own safety and the safety of other coworkers, which may not be considered a core aspect of performance, and thus not receiving appropriate attention. While Katz-Nevon et al. showed the potential utility of safety self-efficacy as a construct, there may be instances in which safety is considered an ancillary behavior that is not directly reinforced by mastery experience, role modeling, or persuasion, as organizations can prioritize productivity and efficiency over safety. It is in fact very likely in organizations prioritizing productivity and profit maximization than workplace safety and health. This being said, safety self-efficacy should still be improved if the individual enacts safety regularly in their job, and the individual should reap the benefits of increased safety self-efficacy but may not be prioritized.

H2b: Experience, determined by set of trials (section), will be positively related to safety self-efficacy.

H3b: Greater levels of safety performance will be related to greater levels of both task and safety self-efficacy, with a greater relationship with safety self-efficacy

There are instances when self-efficacy beliefs can be misinformed, which may have negative repercussions. Self-efficacy beliefs are self-appraisals, which are subject to systematic inaccuracies. Judgments of performance often need structure in order to be accurate and avoid bias (Gatewood, Field & Barrick, 2008). This is especially true when self-appraisals are being made, as there tends to be a leniency bias when someone rates their own performance as opposed to being rated by a third party (Heidemeier & Moser, 2009). Also, research suggests that trait-like self-efficacy exists as a culmination of self-efficacy appraisals over time that is resilient to change (Eden & Aviram, 1993; Shelton, 1990; Sherer et al. 1982). This means that higher general self-efficacy will follow an individual when they make future appraisals on their task self-efficacy. General self-efficacy has been shown to have a positive relationship with task self-efficacy, as well as having a spill-over effect, such that higher general self-efficacy can lead to increases in task self-efficacy across all circumstances (Eden & Aviram, 1993, Shelton, 1990; Sherer et al, 1998). Those that have accumulated a high general self-efficacy, through constant high evaluations of task self-efficacy over time, will often report increased task self-efficacy, establishing a cycle that can result from continued high self-evaluations that may not reflect reality. Likewise, ancillary behaviors, such as workplace safety may be conflated with job performance, especially when the perceived similarity between the job tasks and prior experience is high, leading individuals to believe they are efficacious in their ability to engage in safety, when they are unrelatedly performing well in their job.

H4: Distinct constructs, task and safety self-efficacies will be positively related to one another; while the task self-efficacy to safety self-efficacy relationship will be stronger than safety self-efficacy to task self-efficacy relationship.

In sum, the present study attempted to observe the effect of experience and self-efficacy on task and safety performance, the shifts in task and safety self-efficacy as experience is accrued, and the perception of task and safety self-efficacy. Experience, particularly measured in the number of times a task is completed, would results in improved performance on both task and safety. The repeated task would then allow for appropriate alteration in the individual's self-efficacy for both the task and safety behaviors. Being that safety may not be considered a focal point of the task, the effects of experience and self-efficacy on matters of safety should be weaker than the effects on task performance. Task self-efficacy may also be related to safety self-efficacy in such a way that it can inflate an individual's belief in their ability to remain safe on the job. To this end, an experiment was designed to observe experience, as in number of times a task in completed, and its relationship with subsequent performance and self-efficacy beliefs on a task involving a safety component.

Chapter 2 - Method

Design

This study used a repeated-measure experimental design, in which participants were asked to play multiple trials of the video game “Minesweeper”. Minesweeper is a single-player puzzle game with the objective to clear the game board without detonating one of the hidden “mines”. The game is won when all mine-free squares are revealed, because all mines have been safely avoided. Players are presented a grid containing uniform tiles that have been randomly assigned a set number of mines, and players reveal what is behind the tiles by clicking them. If a mine is present, the tile will turn red and a mine icon will be displayed, indicating that the player lost. If a mine is not present, then a number is displayed in place of the tile, indicating how many mines are adjacent to that tile in all directions. If a tile has no mines adjacent to it, it will be blank. All other adjacent “blank” tiles will be revealed from the original clicked tile until numbered tiles are revealed. Players are tasked with using the numbers as clues to where the mines are, and will continue to navigate the board until they click on a mine or reveal all safe tiles, resulting in a win. The first tile to be clicked per attempt will never reveal a mine, meaning a player will never lose on their first turn. Minesweeper also has a “flag” feature that will mark a tile with a small flag icon and not allow that tile to be interacted with. This allows players to track tiles that they think are dangerous and safeguard themselves from accidentally clicking it, potentially stepping on a mine and losing. The grid in this case will be an 8x8 with 8 mines.

The minesweeper task was modeled after procedures done to simulate a work environment by Probst (2002) in her evaluation of job security and performance quality. Participants were tasked with painting the highest quality images as possible while adhering to basic safety procedures. Each participant took part in a two-hour trial in which they were asked to paint various images that were rated on quality. This quality would determine the “payment” each worker would receive. The safety procedures included rinsing brushes, cleaning splatters, recapping paint, and wearing a safety mask. This procedure provided discrete trials (number of paintings finished), a means of assessing performance (rater coding of painting quality), objective measures of safety behaviors that are mostly separate from performance, and some measure of ecological validity by being incentivized by payment. While this procedure is useful, it comes with disadvantages such as objective rating of performance, hefty time allocation per participant, and physical resources such as space, supplies, and funds to offer participants with the necessary equipment (canvas, paint, brushes, goggles, etc.). Having to wait for a rater to appraise their painting and then grant payment also hampers the adjustment of self-efficacy beliefs per trial. This procedure does not offer readily available objective information that participants can use to appraise their own performance and adjust their self-efficacy.

The minesweeper task attempts to capture a similar level of validity to actual work while providing a streamlined approach to avoid some of the pitfalls of Probst’s procedure. Minesweeper allow for greater levels of experience by having a shorter time of completion per instance than painting. The measure of performance in the minesweeper task is objective rather than subjective, as task performance can be measured in the number of tiles cleared. This also allows for intrinsic feedback within the task, as participants will be able to see their own progress per trial, rather than wait for a rater to judge their painting and then pay them accordingly. This grants participants an opportunity to make judgments about their own performance that are necessary for making adjustments to self-efficacy, a focal variable for this experiment. More importantly, minesweeper also provides an objective measure of safety behavior that is tangentially related to task performance in the form of flag placements. While useful, the flag placement is not necessary for performing well. A player can ignore flags and win, as well as use flags and lose. Finally, performance can still be incentivized. In this case, all participants were entered into a raffle for a $20 amazon gift card, with winning game boards increasing the participants' odds of winning the raffle. Minesweeper is a simple task that can be digested easily and performed quickly, allowing for a streamlined approach to measuring experience, performance, and safety.

Participants and Procedure

78 undergraduate participants were recruited via K-state Sona Systems. Five participants were removed due to administration error, five were removed for careless responding determined by completion of all 45 trials and all survey materials in under 10 minutes, and four removed for having over nine missing values in their survey responses, resulting in 64 participants with appropriate data. The data from survey materials and the associated task was collected anonymously, and no identifying information was obtained. In return for taking part in the survey, participants received class credit and an opportunity to be drawn as a winner in a raffle for five $20 Amazon gift cards, with greater performance allowing for more chances at winning to incentivize greater performance. The mean age of the sample was 19.8, with 67% of the participants (n = 43) identifying as female.

After attaining informed consent, participants responded to a survey portion assessing their general self-efficacy and their prior experience with games including minesweeper. The research assistant then requested participants to watch a 2-minute long video tutorial of minesweeper, detailing visually and auditorily the goal of minesweeper, how to clear tiles, and how to use flags to safely avoid mines. Examples of winning and losing game boards with and without flags were displayed to the participants as well, so they would be privy to the intrinsic feedback that a completed minesweeper game board would provide. Raffle information was then provided both in text via the survey as well as by the research assistant. Participants were then told that they can increase their chances of winning in the raffle by achieving wins in minesweeper. A research assistant let the participants know that their goal is to play a total of 45 trials split into three sections of 15 trials, with the goal to attain as many cleared boards as possible. After each section, a research assistant stopped the participants and directed their attention to items assessing their task and safety self-efficacy. After the final section, demographic information was obtained, and participants were debriefed. Participants' performance was recorded using screen-sharing software and was coded afterward.

Individual Differences as Control Variables

Demographic information and individual differences that are potentially confounded with self-efficacy or performance measures in the present study were collected and used as control variables. These were collected after participants have completed their 45 trials of Minesweeper. These variables include age, sex, prior experience with games including minesweeper, and general self-efficacy. Age and sex were necessary to control for since differences driven by age or sex might be present, such as dexterity with a mouse, exposure to minesweeper, experience with similar video games, and general familiarity with technology. Since general self-efficacy can serve as a baseline for other self-efficacy appraisals (Chen, Gully, & Eden, 2001), it was controlled for at the individual level.

Prior experience

Prior experience was assessed by three self-report 5-point Likert scale (1 = strongly disagree; 5 = strongly agree) items of varying specificity to the task. This variable may not be integral to the question being asked since participants are only being compared to themselves over trial; however, this can serve as a useful variable to help identify any individuals that may come from a different population, such as expert minesweeper players. The first item assessed participants' overall experience with video games ("I have experience playing video games"). The second assessed their experience with puzzle games ("I have experience playing puzzle games"). The final question assessed their prior experience with Minesweeper specifically ("I have experience playing Minesweeper").

Experience accrued during experiment

Experience was organically accrued by participants by the nature of the repeated measures design. Quinones et al. (1995) suggest that the most descriptive and valid form of experience would be measured at the "amount" measurement mode and the "task" level of specificity. Using trial count as a discrete measure of experience satisfies this suggestion while standardizing experience across participants. Experience was also used at the section level, with 15 trials comprising a section, to a total of three sections. This was used so level 2 variables (self-efficacy) could be used in the model along with experience.

General self-efficacy

General self-efficacy (GSE) concerns itself with the trait-like belief in one's overall competence and ability to perform (Chen, Gully, & Eden, 2001). This construct comes with similar work-related benefits such as self-esteem, mastery orientation, and increased motivation (Judge et al., 2000; Chen, Gully, Eden, 2001). When gaps in experience exist and an individual cannot make an appropriate self-appraisal, general self-efficacy can be used to fill these gaps. It is necessary to control for this variable and to confirm that the task and safety self-efficacies are distinct from each other as well as general self-efficacy.

GSE was assessed by Schwazer and Jerusalem's (1995) 10-item 4-point General Self-Efficacy Scale (1 = hardly true; 4 = exactly true). Example items include, "Thanks to my resourcefulness, I know how to handle unforeseen situations" and "I can solve most problems if I invest the necessary effort." These individual responses are then added to form composite scores ranging from 10 to 40. Scholz, Gutiérrez-Doña, Shonali, and Schwarcer (2002) demonstrate the high reliability of this measure ($\alpha = .86$), unidimensionality, and construct validity across 25 countries ($N = 19{,}120$). Validity of the scale has been displayed, as the composite scores did not significantly correlate with age, correlated negatively with a host of negative mental health constructs (e. g. anxiety, depression), and correlated positively with positive mental health constructs (e. g. optimism, social support). Other previous research demonstrates internal consistency and construct validity of the scale (Luszczynska, Gutiérrez-Doña, & Schwazer, 2005; Luszczynska, Scholz, & Schwazer, 2005)

Measures of Focal Variables

Task and safety self-efficacy

In order to assess task and safety self-efficacy for minesweeper game, a series of scale development procedures including candidate item development (five items per each) based on the context self-efficacy literature review (Bandura, 2006) and discussion with subject matter experts (i.e., an industrial/organizational psychology Ph.D. holder, two psychology graduate students, two undergraduate psychology research assistants), exploratory factor analysis (EFA), inter-item correlation and reliability analyses, and confirmatory factor analysis (CFA) were utilized. These procedures returned two two-item scales for assessing task- and safety self-efficacy. Results of these procedures are presented in the result section and the final items respectively for task-and safety self-efficacy scales are presented in the appendix. Example items for task self-efficacy scale include “I am confident that I can become better at Minesweeper.” and “I am confident that I can clear the whole board without triggering a mine.” Example items for safety self-efficacy include “I am confident in my ability to use the flag function to prevent triggering mines.” and “I am confident in my ability to remain safe from mines.” All items were based on 0-100 continuous ratings (appendix B).

Scores for the finalized scales were aggregated, leading to a maximum task self-efficacy score of 200 ($\alpha = .87$; .89; .90 per section respectively), and a safety self-efficacy score of 200 ($\alpha = .90$; .90; .90 per section respectively). These were assessed at three time points; after the initial 15 trials (1st section), after the next 15 trials (2nd section), and after the next 15 trials (3rd section). These items are in accordance with Bandura’s instructions for constructing self-efficacy scales (2006). These self-efficacy scales were developed to gain information on the degrees of assurance an individual has in their ability, not their willingness, by utilizing specific vocabulary such as “can” and “confidence”. Also, it is recommended to supply statements of varying efficacy while allowing individuals to indicate their efficacy beliefs with a broad response scale. Regarding the safety self-efficacy, similar strategies used by Katz-Navon, Naveh, and Stern (2007) were employed by placing emphasis on the efficacy regarding specific safety behaviors.

Task Performance

Task performance was rated by the number of tiles successfully cleared before completion of the individual trial, either by clicking a mine (loss) or by identifying all mines (win). To avoid this task being too difficult, the board has been modified from its original easy setting (10x10 grid with ten mines) to a more manageable board with less mines (8x8 grid with eight mines). Tiles cleared ranged from 1 (the protection from instant loses prevents this being 0) to 56 (a win). The mean number of tiles cleared across all trials was 35.80, with a standard deviation of 19.00.

Safety performance

Safety performance was indicated by engaging in safety behaviors. In this case, it was operationalized as the number of flags used per trial. The accuracy of flag use (properly indicating a bomb with a flag) as well as the misuse (incorrectly marking a tile as a bomb) were both considered safety performance. If the attempt at clearing the board has failed, but the participant still engaged in flag use, these uses will count toward safety performance as well. This is in accordance with operationalization suggested by Griffin and Neal, which break down all job performance into task and contextual performances (Griffin & Neal, 2000). Safety is considered a contextual performance, which may then be further broken down into safety compliance (following recommended safety behaviors) and safety participation (engaging in safety on a cultural level). Measuring only accurate flag use would be remiss, as it would only allow safety performance to be enacted by those capable of identifying mines, which may be an indication of task performance. Allowing all flag use to be considered safety performance ensures that participants safety behaviors are not only measured after attaining mastery. Flags used in a single trial ranged from 0 to 21. The mean number of flags used across all trials was 2.52, with a standard deviation of 3.29.

Analyses

Before hypothesis testing, adequacy of the task and safety self-efficacy measures was assessed. Exploratory factor analysis (EFA) was conducted on the task and safety self-efficacy scales to determine whether the items are assessing the singular proposed construct. An oblique rotation was used, allowing the factors to correlate with each other. Then, three distinct constructs of task (5 initial items; 2 items after trimming), safety (5 initial items; 2 items after trimming), and general self-efficacy (10 items) were specified in a confirmatory factor analysis (CFA) model. This process was to demonstrate the conceptual discrimination across the three distinct self-efficacy constructs in order to proceed with hypothesis testing (discriminant validity). These factor analyses were conducted in the Jamovi software, which is an open source statistical software that utilizes base r packages (The Jamovi Project, 2019).

Multilevel analyses (hierarchical linear modeling) was used to link the within-individual level variables (level 2; task and safety self-efficacy nested within three time points [sections]) with repeated responses from individuals (level 1; experience via trial, task performance per trial, safety performance per trial). Subject characteristics such as age, sex, prior experience with games, as well as general self-efficacy were controlled as level 3 variables. H1a was tested by having the number of tiles cleared (task performance; level 1) regressed on by trial (task experience; level 1). H1b was tested by having the number of flags used (safety performance; level 1) regressed on by trial (task experience). H2a was tested by having the task self-efficacy obtained at the end of each section (level 2) regressed on by section (task experience via section; 3 sections with 15 trials each, level 2). H2a was tested by having the safety self-efficacy obtained at the end of each section (level 2) regressed on by section (task experience via section; 3 sections with 15 trials each, level 2). The level of analysis for hypotheses 2a and 2b was section (level2), because the independent variables were defined at the section level, aggregating them across the 15 trials belonging to each section. By aggregating trials to form sections, the level of analysis was aligned to examine the experience to self-efficacy relationship. H3a and H3b were tested by having task and safety performances regressed on by task self-efficacy and safety self-efficacy in separate models. H4 was tested by having task self-efficacy regressed on by safety self-efficacy, and vice versa. The analyses were conducted using R (R core team, 2016) and ‘lme4’ package (Bates, Mächler, Bolker, & Walker, 2015).

Chapter 3 - Results

Before hypothesis testing, the adequacy of all self-efficacy measures was assessed. Inter-item correlations, reliability analysis, and exploratory factor analyses (EFAs) for both initial 5-item scales for assessing task and safety self-efficacy were conducted to confirm the unidimensionality of the constructs. These were conducted three times per construct to observe similarities across the three sections. The reliabilities of the full five-item task self-efficacy scales at each of the time-points were .85, .89, and .89 respectively. Item means, standard deviations, inter-item correlations, and reliabilities are presented in tables 5-10. Inter-item correlations for the task self-efficacy scale were significant at the $p < .01$ across all time-points. EFA was conducted on the five task self-efficacy items for each time point to observe their factor structure (Tables 11 and 12). While all items loaded onto one factor, items 3 and 4 had the strongest loadings. The reliabilities of the five-item safety self-efficacy scales at each of the time-points were 0.91, 0.91, and 0.92 respectively. Item means, standard deviations, and inter-item correlations are also presented in tables 5-10. EFA was conducted on the five safety self-efficacy scales for each timepoint as well (Tables 13 and 14). Similar to its task counterpart, all of the safety items loaded onto one factor, with two items (1 and 4) consistently having the strongest loading across all three time points.

To ensure that these constructs are discriminating, confirmatory factor analysis was conducted with all pairs of time points. The General Self-Efficacy (GSE) scale was placed into a confirmatory factor analysis (CFA) model with task and safety self-efficacy items for all three sections, generating a total of three CFA models (Tables 16, 17, and 18). Comparative fit index (CFI) and Tucker-Lewis Index (TLI) fit indices indicate strong to moderate fit (i.e., > .95 Hu & Bentler, 1999); however, the root mean square error of approximation (RMSEA) indicates poor fit (i.e., < .08; Hu & Bentler, 1999). For an alternative measurement model, all items besides the strongest loading items for both the task and safety self-efficacies were removed (Tables 19, 20, and 21). The weakest loading item from both scales was removed and a new CFA model was conducted until, all fit indices indicate adequate fit. Task self-efficacy is now being measured by only items 4 and 5, and safety self-efficacy is only being measured with items 1 and 4, creating a potential maximum score of 200. The new means and standard deviations for the final measurements of task and safety self-efficacies are displayed in table 22.

The assumptions of linearity, normality of residuals, and homogeneity of variance, were assessed. Normality was assessed via qq plots per each model, which indicated adequate normality. Levene's tests were used on each model to test homogeneity of variance, in which no model's Levene's test was significant. A plot of residuals by fitted indicated normality of residuals except for models in which the dependent variable was task performance (tiles cleared) or flags used (safety performance) due to the ceiling and floor effects of those variables.

Hypothesis 1 Testing Results

Multilevel modeling showed that participants cleared 34.58, 36.45, and 37.35 tiles on average per section, with 36.06 (0.87) average tiles across all sections. Trial was significantly associated with task performance ($B = .07$ $SE = 3 \times 10^{-3}$, $p < .05$), indicating that an increase in one trial would be associated with a .072 increase in tiles cleared, supporting H1a. Participants used 1.92, 2.81, and 2.83 flags on average per section, with 2.52 (0.18) average flags across all sections. Trial was also significantly related to safety performance ($B = .03$, $SE = 4 \times 10^{-3}$, $p < .01$), such that an increase in trial is associated with a .03 increase in flag use. While these effects are relatively small, their impact over multiple trials can be compounded over time. In order to address H1b, standardized regression coefficients were computed. They indicated that trial's effect on task (*β* = .05, SE = .02) was weaker than trial's effect on safety (*β* = .10, SE = .02), not supporting H1b. Figure 5 and 6 display the average task and safety performance per trial.

Hypothesis 2 Testing Results

Multilevel modeling displayed that participants average task self-efficacy was 102, 116, and 121 per section respectively, with an average of 113 (3.36) across all sections. Task experience coded at the level 2 (section) was significantly associated with task self-efficacy (B = 9.67, SE = .50, p < .01), meaning that a task self-efficacy increases by 9.67 points (on the final 200 point scale) after each section, supporting H2a. Participants average safety self-efficacy was 97, 112, and 117 per section respectively, with an average of 109 (6.79) across all sections. Task experience coded at the level 2 (section) shared a significant relationship with safety self-efficacy (B = 10.18, SE = .44, p < .01); however, the initial model failed to converge. After the four participants that displayed no variance across sections in their self-efficacy were removed from the sample, the model converged. The estimated coefficients were nearly identical to those from the initial model. It can be concluded that safety self-efficacy increased by 10.18 points (on the final 200-point scale) after each section on average, supporting H2b. Figure 7 and 8 display the average task and safety self-efficacy per section.

Hypothesis 3 Testing Results

The initial model testing the effect of performance on task self-efficacy failed to converge due to invariance in task-self efficacy in eight individuals. The model converged after these participants were removed. In support of H3a, both task performance (B = 1.12, SE = .11, p < .01) and safety performance (B = 2.49, SE = .26, p < .01) were significantly and positively related to beliefs of task self-efficacy, with a stronger effect relationship existing between task performance and task self-efficacy. Regarding safety self-efficacy, there was a significant negative relationship with task performance (B = -0.29, SE = .08, p < .01) and a significant positive relationship with safety performance (B = 2.88, SE = .21, p < .01). Figures 9 and 10 illustrate these relationships at section level.

Hypothesis 4 Testing Results

Lastly, task and safety self-efficacy shared significant relationships both when task self-efficacy was regressed on by safety self-efficacy (B = 5.42, SE = .51, p < .01), and when safety self-efficacy was regressed on by task self-efficacy (B = 7.17, SE = .44, p < .01). This supports H4, suggesting that these distinct self-efficacy constructs are interrelated while task self-efficacy having a greater influence on safety self-efficacy.

Chapter 4 - Discussion

Experience is often synonymous with good practice and expertise. While in most cases, experience yields benefit to the individual (Hunter & Hunter,1984; Schmidt, Hunter, & Outerbridge, 1986; McDaniel, Schmidt, & Hunter, 1988), there are cases in which experience can produce maladaptive patterns of behavior (Bilalić, McLeod, & Gobet, 2008; Dokko et al., 2009). Experience can also inform self-efficacy beliefs of individuals about their ability; however, these beliefs might be holistic and not account for different aspects of a task, especially aspects specific to safety that may have no bearing on performance. To assess these assumptions within the domain of safety, an experiment was conducted using minesweeper as a proxy for precarious work. Participants were exposed to 45 trials (3 sections of 15 trials) of minesweeper, with their task performance being measured by the number of tiles cleared and their safety performance being measured by the number of flags used to avoid triggering mines. Self-efficacy beliefs toward their ability to clear as many tiles as possible (task self-efficacy) and their ability to use flags (safety self-efficacy) were assessed after 15 trial intervals (labeled section 1 through 3). Experience in terms of playing more trials was associated with increased task performance by way of clearing more tiles, as well as increased safety performance indicated by increased flag use. Experience in terms of progressing to the next section (playing a chunk of 15 trials) was associated with increases in both task and safety self-efficacy. Performance on the task in previous chunks is positively related to task self-efficacy, but has a negative relationship with safety self-efficacy. Safety performance shares a positive relationship with both task and safety self-efficacies. Both task and safety self-efficacies are related to each other; however, task self-efficacy has greater influence over safety self-efficacy.

In agreement with precedent, experience can yield a positive effect on task performance (Hunter & Hunter, 1984; Schmidt, Hunter, & Outerbridge, 1986; McDaniel, Schmidt, & Hunter, 1988). Meanwhile, findings of the present study suggest that this can also be applied to the performance of safety behaviors as well. This relationship is expected being that experience was measured at the task level of specificity (Quinones et al., 1995), in which the relationship between experience and performance is strongest. Minesweeper requires various cognitive skills such as critical thinking, logical inference, and numeracy as well as motor sensory skills such as visual detection and fine motor coordination for successful performance. These attributes in a given context can be advanced through repeated experiences and accompanying elaboration of a mental model for task performance (Kolz, McFarland, Silverman, 1998; Stajkovik & Luthans, 2003; Sturman, 2003). Minesweeper players were likely to become familiar with the patterns of problems (e.g., which situation is more/less difficult), action-outcome prediction (e.g., clicking a tile will trigger a bomb or show a number in it), best available problem solving approach (e.g., clicking adjacent or distal tile, guessing), and how to maneuver mouse to facilitate accurate clicking behavior (e.g., right vs. left click) throughout a series of trials and errors.

These findings corroborate past literature explaining how experience can inform self-efficacy beliefs. Through the course of salient and adequate experiences in a given task that entail trial and error, sense of mastery is likely to incur. Subsequently, perception on one’s ability to be successful in the task is subject to grow, leading to enhanced self-efficacy. As per Social Cognitive Theory, instances when individuals engage in a task is the primary source for the information necessary to appraise their own performance (Bandura, 1997) as well as the strongest predictor of self-efficacy (Schunk, 1981;1987). By repeatedly offering this mastery experience, individuals are given the opportunity to adjust their self-efficacy and track their improvement. Utilizing the task level of specificity and measuring experience as an amount via trial allows for the mastery experience to be a discrete number. While it is not feasible to measure the fluctuations in self-efficacy beliefs per trial of minesweeper, being able to aggregate to sections that are relatively short (both in terms of the number of trials and time) displays the malleability of self-efficacy beliefs when mastery experience is isolated. The present study took the advantage of a controlled experimental setting to showcase the importance of molding self-efficacy beliefs over time as well as the potential of mastery experience.

Both the factor analyses and hypothesis testing suggest that general self-efficacy (GSE), task self-efficacy, and a safety self-efficacy construct aimed at the performance and compliance of safety behavior, are being perceived differently. These findings corroborate both research suggesting the self-efficacy typically refers to a malleable appraisal (Bandura, 1997), and overall general self-efficacy trait-like construct exists as a culmination of pervious self-efficacy appraisals (Eden & Aviram, 1993; Shelton, 1990; Sherer et al. 1982). In the present study, safety self-efficacy was based on a distinct component of the task, also discriminates from the other two self-efficacy measures, displaying that particular aspects of a task can be understood as different and specific to safety by individuals. While this is the case, these findings suggest that task and safety self-efficacy can inform one another, meaning that shifts in one may spill over to the other (Eden, & Aviram, 1993; Judge, & Bono, 2001).

This is particularly interesting considering that findings also display that task performance shares a negative relationship with safety self-efficacy, as well as the considerable positive relationship between task- and safety self-efficacy, with a greater magnitude of the association from task self-efficacy to safety self-efficacy. These results suggest that participants who held greater perceptions of safety self-efficacy may have been preoccupied with safety, negatively impacting their task performance. It also suggests that participants who showed greater task performance were likely to develop an assumption that their greater task performance might compromise their ability or willingness to perform the task more safely. Also, it can be inferred that high task performers may begin perceiving over the course of their experiences that greater performance is achievable without paying full attention to safety component of the task, learning that safety component is helpful but not necessarily required for task performance. Yet, improved task self-efficacy, which can be a result of better task performance through experience, may be associated with inflated safety self-efficacy as a result self-efficacy constructs spilling over into another. Further research on the safety-self efficacy construct should be conducted to note if its cultivation and expression differs from other specific self-efficacies, with particular attention given to the potential spill-over it may receive from task self-efficacy.

While the general understanding is that experience is related to increased task performance, these findings suggest that this can also applies to safety. Having greater exposure to situations that require safety behaviors to be performed may increase the frequency in which the safety behaviors are engaged. Practically, this may be a precarious method of cultivating safety performance; however, those who encounter situations requiring safety are likely to improve in their ability to perform safety behaviors. Using minesweeper as an example, those exposed to more challenging game boards (i.e., boards in which the mines are spread out, leading to less cleared tiles per click) are more exposed to instances when flag use would be more necessary. This more challenging game board can be understood as a more “dangerous” context, and thus more likely to trigger a mine. While these individuals are gaining experience with flag use, they are also at greater risk. Applying this to a real-world setting can display that experience can lead to improvements in safety, but at the cost of frequent and potentially avoidable exposure to risks and hazards.

Practically, cultivating both task and safety performance and self-efficacy is viable if opportunities to engage in both are being presented to the individual with similar frequency. If a task supplies simultaneous opportunities to engage in safety behaviors as task-based behaviors, the experience that an individual is gaining may serve to improve both safety and task performance in tandem. This being said, it is important to assess and observe them separately as they may not always be comparable and rate in which individuals improve in their safety may not be a function of their improvement in the task. In the case of this experiment, tiles cleared and flags used have different possible ranges per trial despite being counted similarly. In a task with greater ecological validity, the measurement and frequency of task performance may differ greatly from that of safety. Returning to the previous example of a painter (Probst, 2002), painting performance can be operationalized via speed and quality, whereas safety performance would be abstaining from improper ladder use, wearing the appropriate safety equipment, etc. The frequency of making brush strokes as opposed to wearing gear and ascending the ladder will also not be commensurate. While both task and safety performance may be increased by experience concurrently, organizations should not assume that both are at the same level of adequacy from experience alone. In sum, task and safety performance and self-efficacy need to be managed in a balanced way to ensure workers attain the unique skills necessary to perform their jobs efficiently and safely.

The findings from the present study suggest that having greater exposure to situations that require safety behaviors to be performed may increase the frequency in which that are engaged; however, task self-efficacy increases may inflate safety self-efficacy in a potentially unwarranted manner. Moreover, if emphasis is placed disproportionately on task performance than safety performance, it can undermine workers' perceptions regarding their ability or willingness to perform the task safely, unless the safety behaviors are directly incorporate into the metrics of task performance. A potential recommendation to address both of these potential issues is to have organizations supply tailored safety training that goes beyond knowledge and compliance, incorporating guided practice. For this, exploratory learning and after-event reviews for both correct and incorrect safety performance can be considered. Exploratory learning after base instruction would serve as guided experience and increases analogical and adaptive transfer after training (Bell & Kozlowski, 2008). This also allows for the use of self-evaluation activities and after-event reviews. Both display a positive relationship with adaptive transfer and self-regulatory activities, which would be desirable for jobs in which safety is a frequent issue that may require greater thought (Bell & Kozlowski, 2008; Ellis & Davidi, 2005). Specifically, after-event reviews that debrief both successful and unsuccessful attempts yield even more transfer that only reviewing successful procedures. Training should have intentional safety components that allow the individuals to engage in the said tasks, serving as safe and isolated instances of mastery experience that can improve both their performance and self-efficacy in the task.

A strength of this study is the implementation of an experimental method using a task with clear indicators for key variables, which could be adapted to future studies looking to investigate safety. An hour per participant was allocated for this study, meaning that all preparation, questionnaires, and trials were completed within 60 minutes. In a short period of time, a broad scope of rich data can be gathered from participants. The task also offers multiple avenues for objective measurement of key variables. In the present study, experience was measured via trial/section, task performance via tiles cleared, and safety performance via flag use. Alternatively, speed in seconds could have been used to measure performance, number of clicks could have been used to measure efficiency, and time spent between trials observing the game board could have been used as an indicator of willingness/attempts to improve. Difficulty can be adjusted by altering the size of the game board and the number of mines that appear per trials. All this data can be collated after the participant has completed the task, so participants' time can be put toward more trials. This method is also cost effective, being that each participant only needs a computer with screen recording software to participate. This task exhibits potential for ecological validity by applying incentives for any metric that suits the researchers' needs. The minesweeper paradigm's customizability and ease of use allows for a task that can be easily applied to future research questions regarding task performance and safety.

Safety is often considered contextual and cultural, and thus researched from that stance (Griffin & Neal, 2000; Mullen & Kelloway, 2009; Sokas et al., 2009). Using a repeated-measures experimental design offers a unique perspective on how individuals perform safety behaviors and establish subsequent self-appraisals on their own. By utilizing an experimental approach, group level variables that are often the focal point of safety research (culture, supervisor affect, dyadic relationships, teams, etc.) are controlled. The only variable being manipulated is time, allowing for experience to be accrued naturally and in measurable increments. Individuals' performance and ability to comprehend intrinsic feedback were given the opportunity to be cultivated without outside influence. Considering that experience is both a primary method of attaining mastery and gathering information necessary to generate accurate self-efficacy appraisals, it is worthy to reexamine experience as it pertains to safety. It is also worthy to observe the individual in order to understand the interplay between constructs that are expected to vary on an individual level. By having a repeated-measures design that allowed for a meaningful number of trials in a reasonable amount of time, individual-level variables could shift in response to experience. This can lead to a greater understanding of the nature of self-efficacy as it pertains to safety, which can benefit and inform future research.

Limitations and Suggestions for Future Study

There are several limitations that need to be considered when interpreting the results of this experiment. First, the increase in flag use, the indicator of safety performance, may have been due to demand characteristics. The intention of the experiment was to capture the natural practice effect that would occur over multiple iterations of the same task in the same individuals; however, by asking about flag use after the first section, participants may have been reminded about the flags and thus felt encouraged to use the flags more than they would have on their own accord. If demand characteristics exist, then the only significant difference across sections would be that the first section would have a fewer flags than both section 2 and 3. If a true practice effect was present, then section 1 would have less than 2 and 3, and 2 would also have significantly less than 3. In order to test for this, a repeated measures ANOVA was conducted to observe if there were any significant differences in average flag use across section. The omnibus ANOVA was significant ($F(2,124)=20.1$, $p < .01$), and probing displayed that the only significant different section was section 1 ($M=1.91$, $SD=1.80$) with fewer flags than section 2 ($M=2.81$, $SD=2.50$; $t=-4.93$, $p < .01$) and section 3 ($M=2.83$, $SD=2.90$; $t=-5.38$, $p < .01$). While individuals were permitted to use as many flags as they deemed useful, there is a potential range restriction due to the number of mines, as in placing 8 flags would be the ceiling if they are trying to use the flags to identify mines. Coupled with losing on the second turn, average flags use is expected to be relatively low per trial. Due to the nature of flag use and the potential of demand characteristics, findings involving flag use should be interpreted with caution. Future research utilizing a task that allows for a greater range of safety behaviors should be used in future studies to avoid this issue.

It is important to emphasize that the operationalization of safety performance used in this experiment might be more akin to safety compliance; however, precedent has used the terms with subtle differences. In this experiment, safety performance was measured as the use of flags regardless of whether the flag was placed on a mine. This allows for sub-optimal patterns of flag use to count toward safety, such as retroactively flagging mines after the game board was nearly complete or “reflagging” the same tile multiple times. If safety performance was measured as correct flag placement, it may have overlapped greatly with task performance, being that only those with mastery of the task would be able to engage in true safety performance. Regarding past operationalizations of safety performance, Katz-Navon et al. consider safe behaviors conducted while providing health care in a nurse population to be safety performance via their construct “patient safety”, measured by a negatively coded questionnaire detecting unsafe practices that may have been enacted (2006). Griffin and Neal break down performance into task and contextual performance, in which safety is considered a contextual performance that is comprised of safety compliance (following safety protocol) and safety participation (participating in safety on a cultural level) (Griffin & Neal, 2000). While calling indiscriminate flag use safety performance may not be negligent, it may be more accurate and in line with previous safety research to refer to it as compliance. Further research distinguishing compliance and accurate use of safety behaviors should be conducted to see accurate enactment of safety behaviors shares similar relationships with experience and self-efficacy as safety compliance.

The sample was restricted to undergraduate college participants that exhibit careless responding, which may restrict the generalizability of these findings to a broader population. Due to varied patterns of response behavior, it was difficult to detect true careless responding in some cases. Obvious careless responders (those with no variance in their responses, and completed the task to quickly) were removed; however, there are instances when an individual would seemingly place more effort into winning a trial after the first click revealed a large portion of the game board. In most cases, the first few clicks still offer an adequate amount of information to navigate the board, but the appearance of clearing a chunk of tiles in one click may have served as an indicator that the board was now easier to clear. There were also instances when individuals would default to arbitrary response patterns, especially regarding opening strategies on a new trial. Some examples include participants drawing crosses, X's, clicking all four corners, or clicking every odd/even tile from left to right. These strategies were unprompted yet still developed in a noticeable number of individuals. These odd responses may have meaning; however, these patterns were not assessed. This may be a potential avenue for future research regarding rigidity that may come from experience, which has been considered the negative component that can come with repeated unchecked experience (Dokko et al., 2009).

Future research should assess the potential different populations of individuals that exhibit patterns in self-efficacy appraisals as well as those that formed rigidities in performance over their multiple trials. Eight total individuals needed to be removed for some models to converge due to their lack of variance in self-efficacy. No other information from these individuals was invariant, and these individuals were of varying self-efficacies (not limited to the floor [0] or ceiling [200] of self-efficacy) and performances, implying that this might not have been careless responding, but their true belief that their self-efficacy was not wavering. Individuals more resistant to alterations in self-efficacy may react differently to experience, such that those that are not incorporating information from experience to their self-efficacy may need more practice in order to reap the benefits of experience. The persistent and rigid patterns of performance emerging from some individuals should also be examined in future research, as these patterns sometimes lead to decreased performance. These patterns are not an inherent detriment to performance, but stubborn use of these patterns (completion of a pattern in lieu of using the information provided from the game board) can negatively impact performance. An individual may begin drawing a cross on the board, which reveals valuable information about how to navigate the rest of the game board quickly. Instead of ceasing their pattern, as it has already led to attaining the necessary information to win, they complete their pattern with little regard towards the information provided from the board. At worst, this accompanies a greater risk of clicking on a mine. At best, this reduces efficiency by taking longer to clear the board. This rigidity in responding may be problematic in terms of tasks that require adaptability such as minesweeper, however, rigidity towards safety protocol might be desirable, such that individuals form a rigid habit of always engaging in safety behaviors (always wearing a helmet before stepping onto a work site). Further research on rigidity, its formation, and its potential effects on performance and safety is necessary to fully understand how experience effects individual's ability to work and remain safe on the job.

References

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*, 179-211.

Ajzen, I. (2002). Perceived behavioral control, self-efficacy, locus of control, and the theory of planned behavior. *Journal of Applied Social Psychology*, *32*(4), 665-683.

Ash, R. A., & Levine, E. L. (1985). Job applicant training and work experience evaluation: An empirical comparison of four methods. *Journal of Applied Psychology*, *70*(3), 572-576

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

Bandura, A. (1989). Human agency in social cognitive theory. *American Psychologist, 44*(9), 1175-1184.

Bandura, A. (2006). Guide for constructing self-efficacy scales. In F. Pajares & T. Urdan (Eds.). *Self-efficacy beliefs of adolescents, 5,* 307-337). Greenwich, CT: Information Age Publishing.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1-48.

Benach, J., & Muntaner, C. (2007). Precarious employment and health: developing a research agenda. *Journal of Epidemiology and Community Health, 61*(4), 276-277

Beyer, J. M., & Hannah, D. R. (2002). Building on the past: Enacting established personal identities in a new work setting. *Organization Science*, *13*(6), 636-652.

Bidwell, M. (2011). Paying more to get less: The effects of external hiring versus internal mobility. *Administrative Science Quarterly, 56*(3), 369-407.

Bilalić, M., McLeod, P., & Gobet, F. (2008). Inflexibility of experts—Reality or myth? Quantifying the Einstellung effect in chess masters. *Cognitive Psychology*, *56*(2), 73-102.

Bilalić, M., McLeod, P., & Gobet, F. (2008). Why good thoughts block better ones: The mechanism of the pernicious Einstellung (set) effect. *Cognition*, *108*(3), 652-661.

Blume, B. D., Ford, J. K., Baldwin, T. T., & Huang, J. L. (2010). Transfer of training: A meta-analytic review. *Journal of Management*, *36*(4), 1065-1105.

Burke, M. J., Sarpy, S. A., Smith-Crowe, K., Chan-Serafin, S., Salvador, R. O., & Islam, G. (2006). Relative effectiveness of worker safety and health training methods. *American Journal of Public Health*, *96*(2), 315-324.

Campion, M. A., Cheraskin, L., & Stevens, M. J. (1994). Career-related antecedents and outcomes of job rotation. *Academy of Management Journal*, *37*(6), 1518-1542.

Castilla, E. J. (2005). Social networks and employee performance in a call center. *American Journal of Sociology*, *110*(5), 1243-1283.

Chatman, J. A., K. A. Jehn. 1994. Assessing the relationship between industry characteristics and organizational culture: How different can you be? *Academy of Management Journal. 37*(3) 522-553.

Chen, G., Gully, S. M., & Eden, D. (2001). Validation of a new general self-efficacy scale. *Organizational Research Methods*, *4*(1), 62-83

Consiglio, C., Borgogni, L., Di Tecco, C., & Schaufeli, W. B. (2016). What makes employees engaged with their work? The role of self-efficacy and employee's perceptions of social context over time. *Career Development International*, *21*(2), 125-143.

DeJoy, D. M., Smith, T. D., & Dyal, M. A. (2017). Safety climate and firefighting: focus group results. *Journal of Safety Research*, *62*, 107-116.

DeRue, D. S., & Wellman, N. (2009). Developing leaders via experience: the role of developmental challenge, learning orientation, and feedback availability. *Journal of Applied Psychology*, *94*(4), 859.

Dragoni, L., Oh, I. S., Vankatwyk, P., & Tesluk, P. E. (2011). Developing executive leaders: The relative contribution of cognitive ability, personality, and the accumulation of work experience in predicting strategic thinking competency. *Personnel Psychology, 64*(4), 829-864.

Dokko, G., Wilk, S. L., & Rothbard, N. P. (2009). Unpacking prior experience: How career history affects job performance. *Organization Science, 20*(1), 51-68.

Dunning, D. (2011). The Dunning–Kruger effect: On being ignorant of one's own ignorance. *In Advances in Experimental Social Psychology, 44*, 247-296. Academic Press.

Eden, D., & Aviram, A. (1993). Self-efficacy training and speed of reemployment: Helping people to help themselves. *Journal of Applied Psychology*, *78,* 352-360.

Fiedler, F.E. (1970). Leadership experience and leader performance: Another hypothesis shot to hell. *Organizational Behavior and Human Performance*, *5,* 1-14.

Ford, J. K., Quiñones, M. A., Sego, D. J., & Sorra, J. S. (1992). Factors affecting the opportunity to perform trained tasks on the job. *Personnel Psychology*, *45*(3), 511-527.

Gatewood, R. D., Field, H. S., & Barrick, M. (2008). Measures of Job Performance. *Human Resource Selection, 6,* 677-687

Gick, M. L., & Holyoak, K. J. (1987). The cognitive basis of knowledge transfer. In Cormier, S. M., & Hagman, J. D., *Transfer of learning: Contemporary research and applications* (9-46). Academic Press.

Gioia, D. A., & Poole, P. P. (1984). Scripts in organizational behavior. *Academy of Management Review*, *9*(3), 449-459.

Gist, M. E., & Mitchell, T. R. (1992). Self-efficacy: A theoretical analysis of its determinants and malleability. *Academy of Management Review*, *17*, 183-211.

Griffin, M. A., & Neal, A. (2000). Perceptions of safety at work: a framework for linking safety climate to safety performance, knowledge, and motivation. *Journal of Occupational Health Psychology*, *5*(3), 347-360.

Gun, R. T. (1993). The role of regulations in the prevention of occupational injury. *Safety Science*, *16*(1), 47-66.

Hackett, G., & Betz, N. E. (1981). A self-efficacy approach to the career development of women. *Journal of Vocational Behavior*, *18*(3), 326-339.

Hansen, L. (1993). Safety management: A call for (r) evolution. *Professional Safety*, *38*(3), 16-21.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1-55.

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, *96*(1), 72-98

The Jamovi Project (2019). *Jamovi* (Version 0.9) [Computer Software]. Retrieved from https://www.jamovi.org

Judge, T. A., & Bono, J. E. (2001). Relationship of core self-evaluations traits—self-esteem, generalized self-efficacy, locus of control, and emotional stability—with job satisfaction and job performance: A meta-analysis. *Journal of Applied Psychology*, *86*(1), 80-92

Judge, T. A., Jackson, C. L., Shaw, J. C., Scott, B. A., & Rich, B. L. (2007). Self-efficacy and work-related performance: The integral role of individual differences. Journal of Applied Psychology, 92(1), 107-127.

Kartam, N. A., Flood, I., & Koushki, P. (2000). Construction safety in Kuwait: issues, procedures, problems, and recommendations. *Safety Science*, *36*(3), 163-184.

Katz-Navon, T., Naveh, E., & Stern, Z. (2007). Safety self-efficacy and safety performance: Potential antecedents and the moderation effect of standardization. *International Journal of Health Care Quality Assurance, 20*(7), 572-584.

Kolz, A. R., McFarland, L. A., & Silverman, S. B. (1998). Cognitive ability and job experience as predictors of work performance. *The Journal of Psychology, 132*(5), 539-548.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*(6), 1121-113

Lewchuk, W., de Wolff, A., King, A., & Polanyi, M. (2003). From job strain to employment strain: Health effects of precarious employment. *Just Labour, 3*, 23-35

Luszczynska, A., Gutiérrez-Doña, B., & Schwarzer, R. (2005). General self-efficacy in various domains of human functioning: Evidence from five countries. *International Journal of Psychology, 40*(2), 80-89.

Luszczynska, A., Scholz, U., & Schwarzer, R. (2005). The general self-efficacy scale: multicultural validation studies. *The Journal of Psychology, 139*(5), 439-457.

Markus, H., & R. B. Zajonc. (1985). The cognitive perspective in social psychology. G. Lindzey, E. Aronson, eds. *Handbook of Social Psychology*, 3rd ed. Knopf, New York.

McCall, B. P. (1990). Occupational matching: A test of sorts. *Journal of Political Economy*, *98*(1), 45-69.

McDaniel, M. A., Schmidt, F. L., & Hunter, J. E. (1988). Job experience correlates of job performance. *Journal of Applied Psychology, 73*(2), 327-330

Medoff, J. L., & Abraham, K. G. (1980). Experience, performance, and earnings. *The Quarterly Journal of Economics*, *95*(4), 703-736.

Mullen, J. E., & Kelloway, E. K. (2009). Safety leadership: A longitudinal study of the effects of transformational leadership on safety outcomes. *Journal of Occupational and Organizational Psychology, 82*, 253-272.

Ng, T. W., & Feldman, D. C. (2008). The relationship of age to ten dimensions of job performance. *Journal of Applied Psychology, 93*(2), 392-423

Ng, T. W., & Feldman, D. C. (2009). Age, work experience, and the psychological contract. *Journal of Organizational Behavior*, *30*(8), 1053-1075.

Novick, L. R. (1990). Representational transfer in problem solving. *Psychological Science, 1*(2), 128-132.

Pajares, F., & Kranzler, J. (1995). Self-efficacy beliefs and general mental ability in mathematical problem-solving. *Contemporary Educational Psychology*, *20*(4), 426-443.

Probst, T. (2002). Layoffs and tradeoffs: Production, quality, and safety demands under the threat of job loss. *Journal of Occupational Health Psychology*, *7*(3), 211.

Quinlan, M., & Sokas, R. K. (2009). Community campaigns, supply chains, and protecting the health and well-being of workers. *American Journal of Public Health, 99*(S3), S538-S546.

Quińones, M. A., Ford, J. K., & Teachout, M. S. (1995). The relationship between work experience and job performance: A conceptual and meta-analytic review. *Personnel Psychology, 48*(4), 887-910

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rennesund, Å. B., & Saksvik, P. Ø. (2010). Work performance norms and organizational efficacy as cross-level effects on the relationship between individual perceptions of self-efficacy, overcommitment, and work-related stress. *European Journal of Work and Organizational Psychology, 19*(6), 629-653.

Rhee, S., Hur, W., & Kim, M. (2017). The relationship of coworker incivility to job performance and the moderating role of self-efficacy and compassion at work: The job demands-resources (JD-R) approach. *Journal of Business and Psychology, 32*(6), 711-726.

Rousseau, D. M., & Libuser, C. (1997). Contingent workers in high risk environments. *California Management Review, 39*(2), 103-123.

Rowe, P. M. (2018). Work experience, the scientist-practitioner model, and cooperative education. *Canadian Psychology, 59*(2), 144-150.

Scholz, U., Gutiérrez-Doña, B., Sud, S., & Schwarzer, R. (2002). Is general self-efficacy a universal construct? Psychometric findings from 25 countries. *European Journal of Psychological Assessment, 18*(3), 242-251

Schmidt, F. L., Hunter, J. E., & Outerbridge, A. N. (1986). Impact of job experience and ability on job knowledge, work sample performance, and supervisory ratings of job performance. *Journal of Applied Psychology, 71*, 432–439

Schunk, D. H. (1981). Modeling and attributional effects on children's achievement: A self-efficacy analysis. *Journal of Educational Psychology*, *73*(1), 93.

Schunk, D. H. (1987). Peer models and children’s behavioral change. *Review of Educational Research*, *57*(2), 149-174.

Schwarzer. R.. & Jerusalem. M. (1995). Optimistic self-beliefs as a resource factor in coping with stress. In S.E. Hobfoll & M.W. deVries (Eds.). *Extreme stress and communities: Impact and intervention* (159-177). Dordrecht, The Netherlands: Kluwer.

Shelton, S. H. (1990). Developing the construct of general self-efficacy. *Psychological Reports*, *66*, 987-994.

Sherer, M., Maddux, J. E., Mercandante, B., Prentice-Dunn, S., Jacobs, B., & Rogers, R. W. (1982). The self-efficacy scale: Construction and validation. *Psychological Reports*, *51*(2), 663-671.

Singley, M. K., & Anderson, J. R. (1989). The transfer of cognitive skill (9). *Harvard University Press.*

Sokas, R. K., Jorgensen, E., Nickels, L., Gao, W., & Gittleman, J. L. (2009). An intervention effectiveness study of hazard awareness training in the construction building trades. *Public Health Reports*, *124*(4), 161-168.

Stajkovic, A. D., & Luthans, F. (2003). Behavioral management and task performance in organizations: conceptual background, meta-analysis, and test of alternative models. *Personnel Psychology*, *56*(1), 155-194.

Sturman, M. C. (2003). Searching for the inverted U-shaped relationship between time and performance: Meta-analyses of the experience/performance, tenure/performance, and age/performance relationships. *Journal of Management*, *29*(5), 609-640.

Tai, W. T. (2006). Effects of training framing, general self-efficacy and training motivation on trainees' training effectiveness. *Personnel Review*, *35*(1), 51-65.

U.S. Department of Labor, Bureau of Labor Statistics. (2017, November 9). *Employer-reported Workplace Injuries and Illnesses -2016.* Retrieved from: https://www.bls.gov/news.release/archives/osh_11092017.pdf

U.S. Department of Labor, Bureau of Labor Statistics. (2017, December 19). *National census of fatal occupational injuries in 2016*. Retrieved from: https://www.bls.gov/news.release/pdf/cfoi.pdf

U. S. Department of Labor Occupational Safety and Health Administration. Training *requirements and resources.* Retrieved from: https://www.osha.gov/dte/

Waldman, D. A., & Avolio, B. J. (1986). A meta-analysis of age differences in job performance. *Journal of Applied Psychology, 71*(1), 33-38.

Walsh, J. P. (1995). Managerial and organizational cognition: Notes from a trip down memory lane. *Organization Science, 6*(3), 280-321.

Woltz, Dan J., Michael K. Gardner, and Brian G. Bell. "Negative transfer errors in sequential cognitive skills: Strong-but-wrong sequence application." *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(3), 601-625.

Yeow, P. H., & Sen, R. N. (2003). Quality, productivity, occupational health and safety and cost effectiveness of ergonomic improvements in the test workstations of an electronic factory. *International Journal of Industrial Ergonomics*, *32*(3), 147-163.

Zimmerman, B. J. (1990). Self-regulating academic learning and achievement: The emergence of a social cognitive perspective. *Educational Psychology Review*, *2*(2), 173-201.

Zohar, D. (1980). Safety climate in industrial organizations: theoretical and applied implications. *Journal of Applied Psychology, 65*, 96–102

Appendix A - Figures

Level of Specificity		Amount	Time	Type
	Organization	Number of organizations	Org. tenure / seniority	Type of organization
	Job	Number of jobs or aggregate number of unique tasks	Job tenure / seniority	Job complexity
	Task	Number of times performing a task	Time on task	Task difficulty
		Measurement Mode		

Figure 1. Specificity and measurement mode of experience

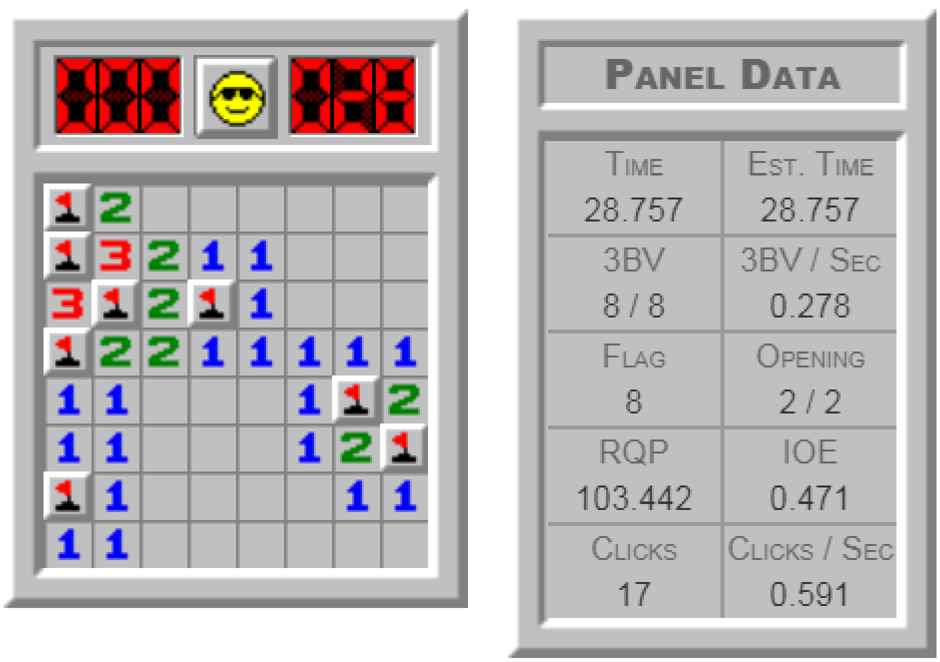

Figure 2. A successfully completed game of 8x8 Minesweeper

	Task		
	Painting Task	Painting Task; Probst (2002) procedure	Minesweeper procedure; present study
Task Performance	· Area (square feet) · Number of jobs completed	· Coder rating (is not free from subjectivity)	· Number of tiles cleared (objective measure)
Safety Performance	· Use of proper equipment · Abstinence from unsafe behavior	· Use of equipment · Cleaning equipment	· Number of flags placed
Incentive	· Payment · Positive feedback from employer	· Payment via quality	· Increased raffle chances · Performance feedback after each trial

Figure 3. Comparison of performance, safety, and incentive across task and safety performance evaluation paradigms

Before task	• Assessment of control variables • Video tutorial • Raffle Information
	Section 1 (Trials 1-15)
After section 1	• Assessment of task self-efficacy • Assessment of safety self-efficacy
	Section 2 (Trials 16-30)
After section 2	• Assessment of task self-efficacy • Assessment of safety self-efficacy
	Section 3 (Trails 31-45)
After section 3	• Assessment of task self-efficacy • Assessment of safety self-efficacy
After task	• Assessment of demographic information

Figure 4. Visualized Experiment Procedure

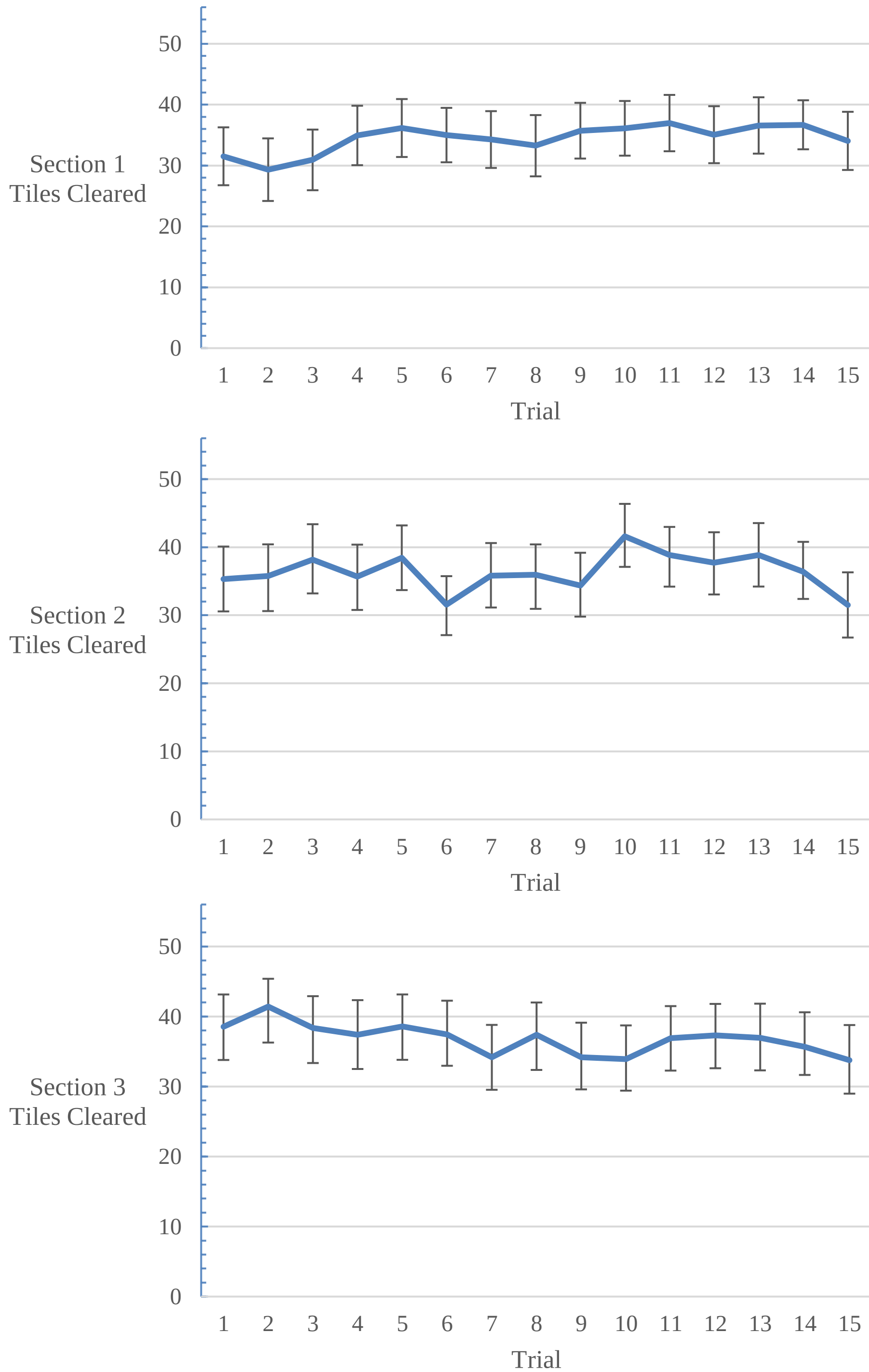

Figure 5. Average number of tiles cleared per trial

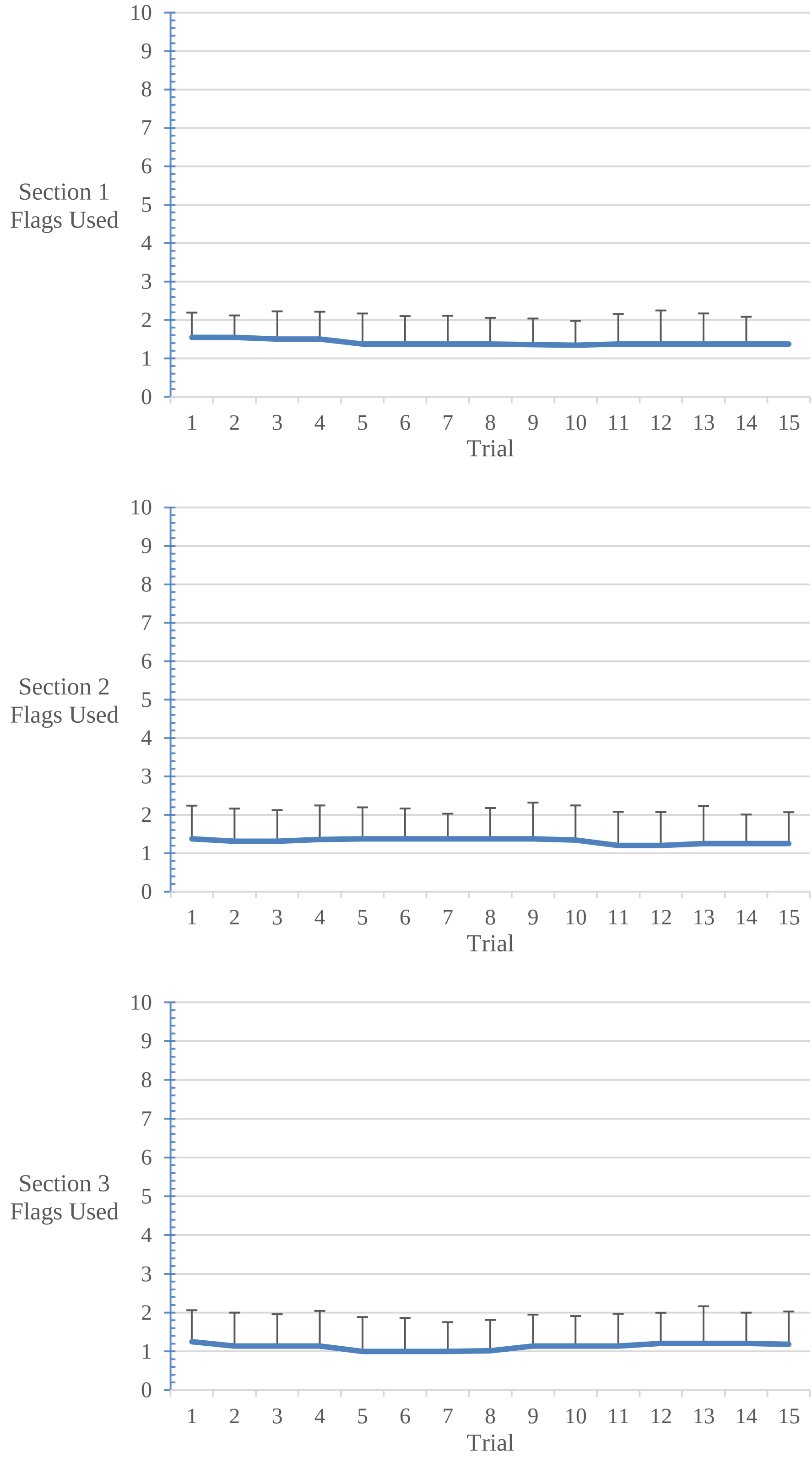

Figure 6. Average number of flags used per trial

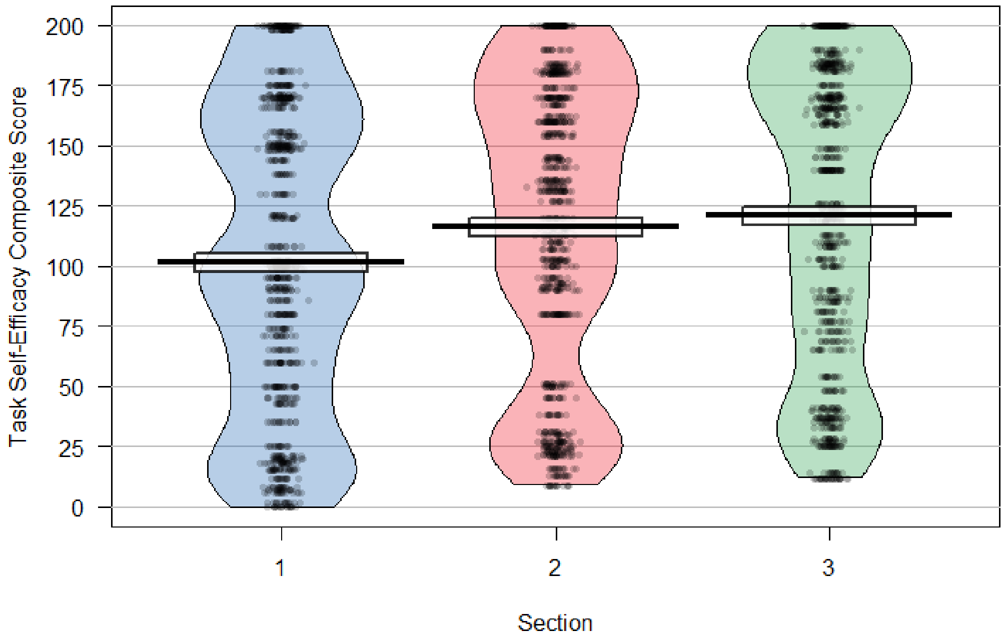

Figure 7. Task self-efficacy across section

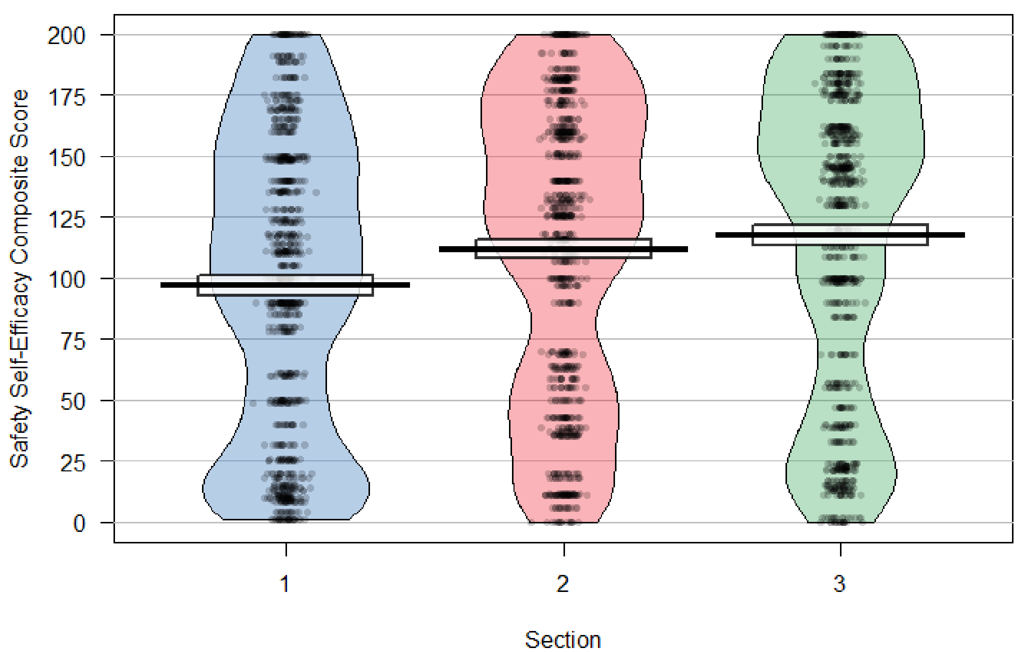

Figure 8. Safety self efficacy across section

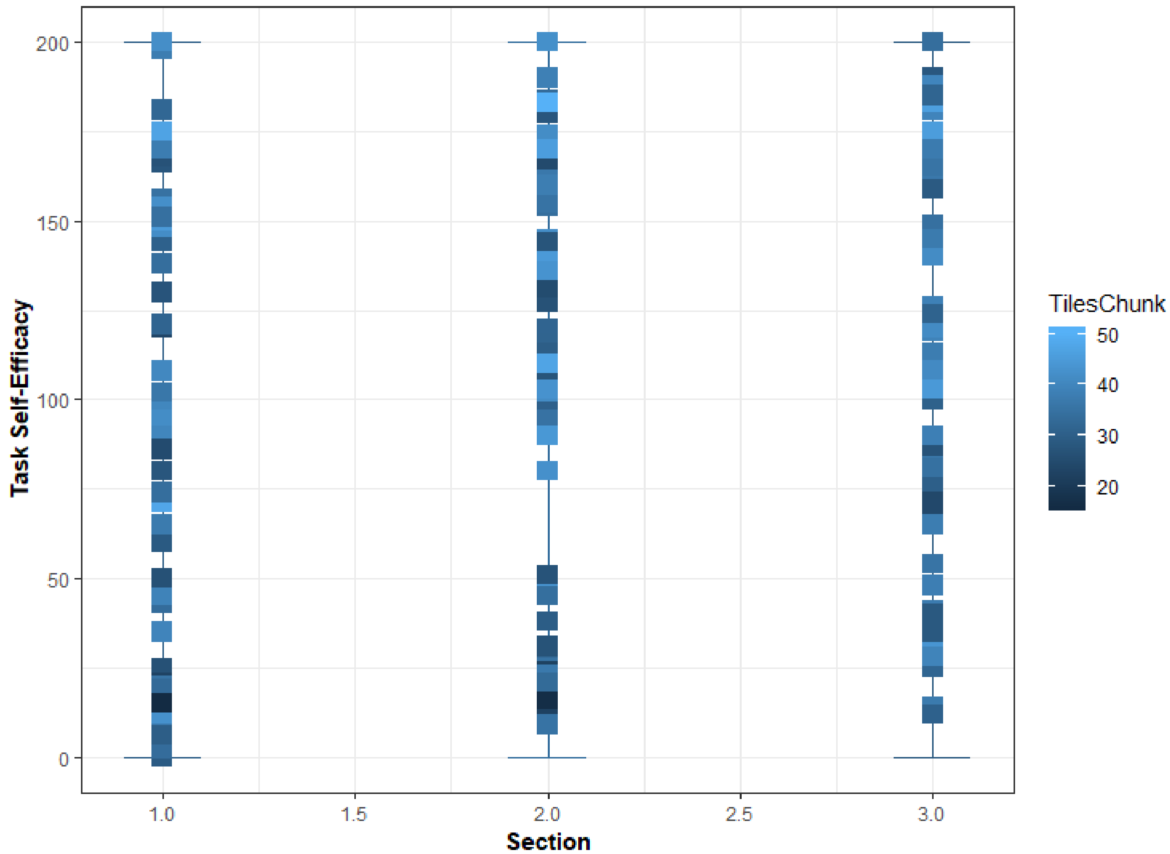

Figure 9. Tiles cleared at the level 2 and task self-efficacy across section

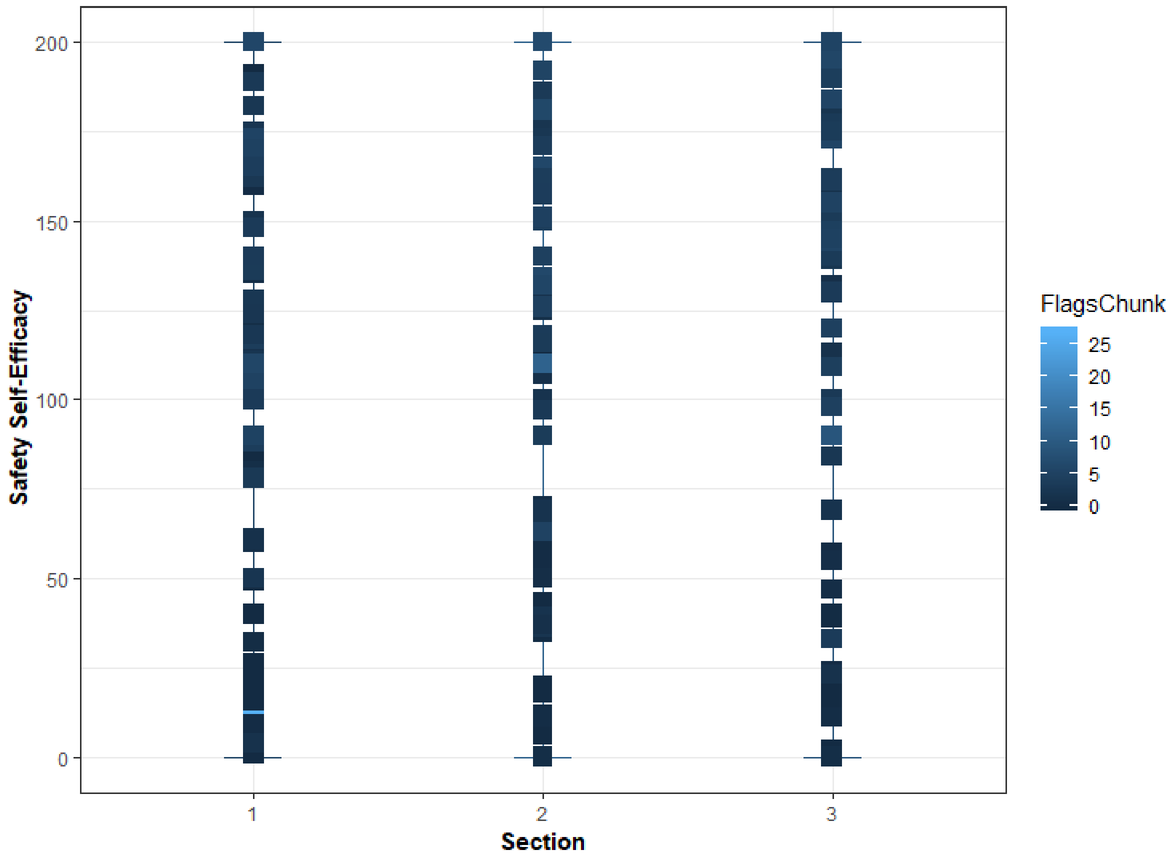

Figure 10. Flags used at the level 2 and safety self-efficacy across section

Appendix B - Tables

Table 1. Item Reliability Statistics for the General Self-Efficacy Scale

	Mean	SD	Cronbach's α if dropped
Item 1	3.25	.50	.89
Item 2	2.73	.54	.89
Item 3	3.22	.68	.89
Item 4	3.03	.76	.87
Item 5	3.08	.76	.88
Item 6	3.42	.61	.88
Item 7	2.95	.77	.88
Item 8	3.00	.74	.88
Item 9	3.25	.67	.87
Item 10	3.23	.58	.88
Full Scale	3.12	.48	.89

Table 2. Correlation Matrix for General Self-Efficacy Items

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10
Item 1	—	.31*	.12	.33*	.44***	.32**	.40**	.56***	.50***	.45***
Item 2		—	.29*	.45***	.28*	.25*	.47***	.32*	.41***	.35**
Item 3			—	.42***	.46***	.50***	.20	.13	.37**	.51**
Item 4				—	.66***	.56***	.55***	.46***	.68***	.52***
Item 5					—	.58***	.44***	.51***	.68***	.53***
Item 6						—	.31*	.49***	.59***	.52***
Item 7							—	.59***	.58***	.45***
Item 8								—	.68***	.48***
Item 9									—	.70***
Item 10										—

Note. * $p < .05$, ** $p < .01$, *** $p < .001$

Table 3. Factor loadings for General Self-Efficacy Scale with one factor

Indicator	Unstandardized	SE	Standardized
Item 1	.28	.06	4.77
Item 2	.26	.07	3.90
Item 3	.32	.08	3.82
Item 4	.57	.08	6.99
Item 5	.58	.08	7.03
Item 6	.41	.07	5.95
Item 7	.50	.09	5.68
Item 8	.52	.08	6.42
Item 9	.60	.07	9.12
Item 10	.43	.06	6.82

Table 4. Model fit for a single factor for GSE

χ^2	*df*	χ^2/df	*p*	CFI	TLI	RMSEA	Lower 90% CI	Upper 90% CI
61.40	35	1.75	< .01	.91	.89	.11	.06	.15

Table 5. Correlation Matrix for Section 1 Task and Safety self-efficacies

	Item 1	Item 2	Item 3	Item 4	Item 5
Item 1	—	.51	.51	.54	.46
Item 2	.81	—	.54	.53	.34
Item 3	.81	.66	—	.77	.56
Item 4	.74	.49	.73	—	.53
Item 5	.61	.73	.56	.49	—

Notes. Above the divide displays correlations for task self-efficacy, below displays safety self-efficacy. All correlations are significant at the p < .001 level.

Table 6. Item Descriptive and Reliability Statistics for Section 1

	Task Self-Efficacy			Safety Self-Efficacy		
	Mean	SD	Cronbach's α if dropped	Mean	SD	Cronbach's α if dropped
Item 1	59.20	24.90	.83	50.30	31.40	.86
Item 2	19.50	24.70	.84	31.80	30.90	.88
Item 3	54.00	33.00	.79	46.70	33.70	.87
Item 4	47.60	34.20	.79	65.10	36.70	.90
Item 5	74.00	26.50	.84	36.50	27.30	.90
Scale	50.90	22.90	.85	46.10	27.40	.91

Table 7. Correlation Matrix for Section 2 Task and Safety self-efficacies

	Item 1	Item 2	Item 3	Item 4	Item 5
Item 1	—	.60	.62	.63	.55
Item 2	.76	—	.57	.63	.52
Item 3	.83	.70	—	.80	.65
Item 4	.73	.55	.79	—	.64
Item 5	.62	.73	.58	.55	—

Note. Above the divide displays correlations for task self-efficacy, below displays safety self-efficacy. All correlations are significant at the $p < .001$ level

Table 8. Item Descriptive and Reliability Statistics for Section 2

	Task Self-Efficacy			Safety Self-Efficacy		
	Mean	SD	Cronbach's α if dropped	Mean	SD	Cronbach's α if dropped
Item 1	67.20	21.60	.88	57.60	31.20	.88
Item 2	33.10	27.90	.88	36.00	32.10	.80
Item 3	63.00	32.00	.85	54.30	32.60	.88
Item 4	53.20	32.40	.84	67.70	34.50	.90
Item 5	75.00	25.90	.87	39.90	26.20	.91
Scale	58.30	23.50	.89	51.10	27.10	.91

Table 9. Correlation Matrix for Section 3 Task and Safety self-efficacies

	Item 1	Item 2	Item 3	Item 4	Item 5
Item 1	—	.59	.49	.50	.47
Item 2	.79	—	.62	.70	.52
Item 3	.87	.69	—	.82	.70
Item 4	.75	.55	.82	—	.67
Item 5	.67	.71	.61	.61	—

Note. Above the divide displays correlations for task self-efficacy, below displays safety self-efficacy. All correlations are significant at the $p < .001$ level

Table 10. Item Descriptive and Reliability Statistics for Section 3

	Task Self-Efficacy			Safety Self-Efficacy		
	Mean	SD	Cronbach's α if dropped	Mean	SD	Cronbach's α if dropped
Item 1	73.50	22.20	.89	61.30	33.50	.89
Item 2	39.30	29.70	.86	40.80	33.50	.91
Item 3	65.40	33.20	.84	56.10	33.00	.89
Item 4	55.60	32.50	.84	65.40	35.00	.91
Item 5	76.00	28.30	.87	44.40	29.50	.92
Scale	62.00	24.40	.87	53.60	28.90	.92

Table 11. Factor Loadings for Task Self-Efficacy

	Section 1	Section 2	Section 3
	Factor 1	Factor 1	Factor 1
Item 1	.67	.75	.62
Item 2	.64	.72	.76
Item 3	.86	.86	.87
Item 4	.85	.89	.90
Item 5	.64	.73	.75
% Explained	54.40	62.70	61.90

Note. 'Principal axis factoring' extraction method was used in combination with an 'oblimin' rotation

Table 12. Model Fit for CFA on Task Self-Efficacy (based on the initial 5-item measure)

Section	RMSEA	Lower CI	Upper CI	TLI	χ^2	df	χ^2/df	p
1	.05	.00	.18	.99	5.51	5	1.10	.36
2	.03	.00	.18	.99	5.05	5	1.01	.41
3	.14	.00	.25	.93	10.09	5	2.18	.05

Table 13. Factor Loadings for Safety Self-Efficacy

	Section 1	Section 2	Section 3
	Factor 1	Factor 1	Factor 1
Item 1	.95	.90	.94
Item 2	-83	.82	.81
Item 3	.86	.89	.91
Item 4	.74	.79	.81
Item 5	.71	.73	.75
% Explained	67.20	68.70	71.40

Note. 'Principal axis factoring' extraction method was used in combination with an 'oblimin' rotation

Table 14. Model Fit for CFA on Safety Self-Efficacy (based on the initial 5-item measure)

Section	RMSEA	Lower CI	Upper CI	TLI	χ^2	df	χ^2/df	*p*
1	.30	.20	.39	.77	31.20	5	6.24	< .01
2	.24	.14	.34	.85	22.20	5	4.44	< .01
3	.26	.16	.36	.84	25.30	5	5.06	< .01

Table 15. Factor loadings for all sections using all indicators

Factor	Item	Section 1			Section 2			Section 3		
		Unstandardized	SE	Standardized	Unstandardized	SE	Standardized	Unstandardized	SE	Standardized
1	GSE Item 1	.29	.06	4.80	.28	.06	4.78	.28	.06	4.78
	GSE Item 2	.26	.07	3.93	.26	.07	3.94	.26	.07	3.96
	GSE Item 3	.32	.08	3.84	.32	.08	3.83	.32	.08	3.85
	GSE Item 4	.58	.08	7.08	.57	.08	7.06	.57	.08	7.05
	GSE Item 5	.57	.08	6.97	.57	.08	6.98	.58	.08	7.01
	GSE Item 6	.41	.07	5.95	.41	.07	5.93	.41	.07	5.91
	GSE Item 7	.50	.09	5.70	.05	.09	5.68	.49	.09	5.67
	GSE Item 8	.52	.08	6.40	.52	.08	6.41	.52	.08	6.37
	GSE Item 9	.60	.07	9.11	.60	.07	9.14	.60	.07	9.14
	GSE Item 10	.43	.06	6.77	.43	.06	6.77	.43	.06	6.80
2	TSE Item 1	15.56	2.88	5.40	15.37	2.38	6.45	12.88	2.59	4.98
	TSE Item 2	17.21	2.80	6.16	20.50	3.05	6.72	22.24	3.20	6.95
	TSE Item 3	27.31	3.45	7.91	27.29	3.24	8.42	29.27	3.27	8.94
	TSE Item 4	27.83	3.62	7.69	28.61	3.20	8.93	29.39	3.15	9.34
	TSE Item 5	18.24	2.99	6.01	19.01	2.82	6.74	20.94	3.06	6.85
3	SSE Item 1	28.79	3.01	9.55	28.10	3.03	9.28	30.81	3.18	9.68
	SSE Item 2	25.44	3.18	7.99	25.80	3.34	7.72	26.71	3.50	7.64
	SSE Item 3	28.50	3.40	8.37	28.76	3.22	8.95	30.06	3.17	9.50
	SSE Item 4	28.19	3.90	7.23	27.42	3.62	7.57	28.75	3.59	8.00
	SSE Item 5	19.92	3.01	6.63	19.40	2.87	6.77	22.02	3.18	6.92

Table 16. Factor Covariances for all sections

		Unstandardized	SE	Standardized
Section 1				
Factor 1	Factor 1	1.00		
	Factor 2	.41	.12	3.44
	Factor 3	.38	.12	3.25
Factor 2	Factor 2	1.00		
	Factor 3	.92	.04	21.68
Section 2				
Factor 1	Factor 1	1.00		
	Factor 2	.37	.12	3.12
	Factor 3	.37	.12	3.13
Factor 2	Factor 2	1.00		
	Factor 3	.87	.04	20.31
Section 3				
Factor 1	Factor 1	1.00		
	Factor 2	.42	.11	3.71
	Factor 3	.33	.12	2.73
Factor 2	Factor 2	1.00		
	Factor 3	.86	.04	19.92

Table 17. Model fit for 3-Factor CFA solution based on all indicators at all sections

Section	χ^2	df	χ^2/df	*p*	CFI	TLI	RMSEA	Lower 90% CI	Upper 90% CI
1	316	167	1.89	< .001	.82	.80	.12	.10	.14
2	292	167	1.75	< .001	.86	.84	.11	.09	.13
3	390	167	2.34	< .001	.78	.75	.15	.13	.16

Table 18. Factor loadings using the two strongest indicators for each self-efficacy measure

Factor	Item	Section 1			Section 2			Section 3		
		Unstandardized	SE	Standardized	Unstandardized	SE	Standardized	Unstandardized	SE	Standardized
1	GSE Item 1	.28	.06	4.79	.28	.06	4.77	.28	.06	4.77
	GSE Item 2	.26	.07	3.93	.26	.07	3.94	.26	.07	3.96
	GSE Item 3	.32	.08	3.84	.32	.08	3.84	.32	.08	3.87
	GSE Item 4	.57	.08	7.06	.57	.08	7.06	.57	.08	7.05
	GSE Item 5	.57	.08	6.97	.57	.08	6.98	.58	.08	7.04
	GSE Item 6	.41	.07	5.96	.41	.07	5.94	.41	.07	5.91
	GSE Item 7	.50	.09	5.69	.50	.09	5.68	.49	.09	5.66
	GSE Item 8	.52	.08	6.42	.52	.08	6.40	.52	.08	6.35
	GSE Item 9	.60	.07	9.12	.60	.07	9.16	.60	.07	9.14
	GSE Item 10	.43	.06	6.77	.43	.06	6.77	.43	.06	6.80
2	TSE Item 3	28.95	3.48	8.33	27.84	3.34	8.34	30.34	3.27	9.27
	TSE Item 4	29.33	3.64	8.06	29.38	3.32	8.85	28.69	3.27	8.79
3	SSE Item 1	29.26	3.15	9.28	28.88	3.14	9.20	29.81	3.33	8.95
	SSE Item 3	28.96	3.52	8.24	28.72	3.36	8.55	31.60	3.14	10.07

Table 19. Factor Covariances for truncated scales

Section 1				
		Estimate	SE	Z
Factor 1	Factor 1	1.00		
	Factor 2	.35	.13	2.78
	Factor 3	.36	.12	2.98
Factor 2	Factor 2	1.00		
	Factor 3	.79	.07	11.63
Section 2				
Factor 1	Factor 1	1.00		
	Factor 2	.36	0.12	2.97
	Factor 3	.37	0.12	3.06
Factor 2	Factor 2	1.00		
	Factor 3	.80	0.06	13.06
Section 2				
Factor 1	Factor 1	1.00		
	Factor 2	.41	.12	3.58
	Factor 3	.29	.12	2.29
Factor 2	Factor 2	1.00		
	Factor 3	.83	.05	15.56

Table 20. Model fit for 3-Factor CFA solution based on truncated scales at all sections

χ^2	df	χ^2/df	p	CFI	TLI	RMSEA	Lower 90% CI	Upper 90% CI
111	74	1.50	$< .01$	.92	.90	.09	.05	.12
101	74	1.36	.02	.95	.93	.08	.03	.11
120	74	1.62	$< .01$	.92	.90	.10	.06	.13

Table 21. Descriptive statistics for final self-efficacy scales

		Section 1			
		Mean	SD	Min	Max
Task					
	Item 3	54.00	33.00	0	100
	Item 4	47.60	34.20	0	100
Safety					
	Item 1	50.30	31.40	0	100
	Item 3	46.70	33.70	0	100
		Section 2			
		Mean	SD	Min	Max
Task					
	Item 3	63.00	32.00	0	100
	Item 4	53.20	32.40	0	100
Safety					
	Item 1	57.60	31.20	0	100
	Item 3	54.30	32.60	0	100
		Section 3			
		Mean	SD	Min	Max
Task					
	Item 3	65.40	33.20	3	100
	Item 4	55.60	32.50	2	100
Safety					
	Item 1	61.30	33.50	0	100
	Item 3	56.10	33.00	0	100

Table 22. Means and standard deviations (values within parentheses) for tiles, flags, task self-efficacy (TSE), and safety self-efficacy (SSE)

	Section 1		Section 2		Section	
Trial	Tiles	Flags	Tiles	Flags	Tiles	Flags
1	31.50 (19.20)	1.55 (2.61)	35.30 (18.90)	2.53 (3.45)	46.00 (64.20)	3.00 (3.29)
2	29.30 (20.80)	1.14 (2.32)	35.80 (21.00)	2.61 (3.28)	42.30 (15.30)	3.89 (3.49)
3	30.90 (20.20)	1.84 (2.93)	37.90 (19.10)	3.08 (3.59)	38.30 (18.50)	3.02 (3.32)
4	34.90 (19.80)	1.91 (2.89)	35.70 (19.20)	2.53 (3.32)	37.40 (20.00)	3.06 (3.66)
5	36.20 (19.30)	2.17 (3.21)	38.50 (17.00)	2.72 (3.20)	38.60 (18.50)	3.08 (3.59)
6	35.00 (18.10)	2.08 (2.94)	32.60 (20.00)	2.11 (2.90)	37.40 (19.50)	2.95 (3.50)
7	33.40 (19.10)	1.92 (2.89)	35.80 (18.00)	2.63 (3.25)	34.20 (18.60)	2.27 (3.06)
8	33.30 (20.40)	1.77 (2.76)	36.00 (19.40)	2.95 (3.82)	37.40 (18.60)	3.02 (3.22)
9	35.70 (18.50)	1.83 (2.75)	34.40 (19.30)	2.75 (3.65)	34.20 (20.00)	2.39 (3.27)
10	36.10 (18.20)	1.64 (2.57)	41.60 (16.80)	3.39 (3.55)	33.90 (19.60)	2.55 (3.13)
11	37.00 (18.70)	2.16 (3.16)	38.80 (18.00)	3.14 (3.52)	36.90 (18.60)	2.53 (3.35)
12	35.10 (19.00)	2.31 (3.54)	37.70 (19.00)	3.25 (3.96)	37.30 (18.30)	2.73 (3.22)
13	36.60 (18.70)	2.36 (3.22)	38.80 (17.70)	2.81 (3.08)	36.90 (19.80)	2.81 (3.91)
14	36.70 (16.30)	1.86 (2.87)	36.40 (19.40)	2.91 (3.31)	35.70 (19.90)	2.56 (3.23)
15	34.00 (19.30)	2.28 (3.50)	31.50 (20.70)	2.80 (3.64)	33.80 (20.30)	2.61 (3.41)
	Section mean		Section mean		Section mean	
Tiles/Flags	34.38 (19.04)	1.92 (2.94)	36.45 (18.90)	2.81 (2.94)	37.35 (21.98)	2.83 (3.38)
TSE	102.00 (63.10)		116.00 (61.10)		121.00 (62.70)	
SSE	97.10 (62.00)		112.00 (61.00)		117.00 (64.30)	
	Mean across all sections					
Tiles/Flags	36.06 (21.14)	2.52 (3.29)				
TSE	113.00 (62.50)					
SSE	109.00 (62.70)					

Notes. "Tiles" represents the number of tiles cleared. "Flags" represents the number of flags used. "TSE" represents the task self-efficacy score on the finalized 200-point scale. "SSE" represents the safety self-efficacy score on the finalized 200-point scale

Appendix C - Scales

Measures of Control Variables

Prior experience to games
5-point Likert Scale: Strongly disagree, Somewhat disagree, Neither agree nor disagree, Somewhat agree, Strongly agree.

1) I have experience playing video games.
2) I have experience playing puzzle games.
3) I have experience playing Minesweeper.

General Self Efficacy
4-point Likert Scale: Not true at all, Hardly true, Moderately true, Exactly true

1) I can always manage to solve difficult problems if I try hard enough.
2) If someone opposes me, I can find the means and ways to get what I want.
3) It is easy for me to stick to my aims and accomplish my goals.
4) I am confident that I could deal efficiently with unexpected events.
5) Thanks to my resourcefulness, I know how to handle unforeseen situations.
6) I can solve most problems if I invest the necessary effort.
7) I can remain calm when facing difficulties because I can rely on my coping abilities.
8) When I am confronted with a problem, I can usually find several solutions.
9) If I am in trouble, I can usually think of a solution.
10) I can usually handle whatever comes my way.

Measures of Self-Efficacy

Task Self-Efficacy

Please rate how certain you are that you can achieve the goals described below.
Rate your degree of confidence by recording a number from 0 to 100 using the scale given below:

Confidence

(Cannot do at all)					(Moderately do)					(Highly certain can do)
0	10	20	30	40	**50**	60	70	80	90	**100**

1) I am confident that I can clear half of the board without triggering a mine.
2) I am confident that I can clear the whole board without triggering a mine.
3) I am confident that I can understand the logic behind the Minesweeper game.
4) I am confident that I can teach others how to play this game correctly.
5) I am confident that I can become better at Minesweeper.

Safety Self-Efficacy

Please rate how certain you are that you can achieve the goals described below.
Rate your degree of confidence by recording a number from 0 to 100 using the scale given below:

Confidence

(Cannot do at all)					(Moderately do)					(Highly certain can do)
0	10	20	30	40	**50**	60	70	80	90	**100**

1) I am confident that I can use half of the flags.
2) I am confident that I can use all of the flags.
3) I am confident in my ability to use the flag function to prevent triggering mines.
4) I am confident that I understand how flag use can help me avoid mines.
5) I am confident in my ability to remain safe from mines.