Technology, Knowledge and Learning
https://doi.org/10.1007/s10758-025-09852-8

ORIGINAL RESEARCH

Calibration Accuracy of General and Task-Specific Learning Self-Efficacy in a Military Training Simulator

Helena J. M. Pennings[1,2] · **Sindu V. George**[3] · **Mathilde E. Meijer**[4,5]

Accepted: 24 April 2025

Abstract

This study contributes to the understanding of the impact of self-efficacy in a high-fidelity simulator training by unpacking the association between general self-efficacy, task-specific self-efficacy, and performance. Miscalibration of self-efficacy among the trainees was studied, analyzing differences in overestimating, accurate, and underestimating trainees in terms of their overall performance. Participants were 66 military trainees who were training to become gunners in a CV90 armored vehicle. Various log-data were used as objective performance indicators. General self-efficacy was measured prior to the training. Task-specific self-efficacy was measured three times: pre-training, mid-training, and post-training. General self-efficacy and overall performance were positively related. About 45–55% of trainees miscalibrated their task-specific self-efficacy to their overall performance (i.e., varying per time-point); the majority of trainees overestimated their capacities. Furthermore, overestimating trainees performed worse than accurate and underestimating trainees, making more *high severity*, *medium severity*, and *low severity* errors in almost all specific tasks, except target detection. Being the first of its kind, our study adds to the existing knowledge base of miscalibration of general learning and task-specific self-efficacy in a military training simulator. Suggestions for further research and implications are discussed.

Keywords Self-efficacy · Self-efficacy bias · Calibration accuracy · Self-evaluation bias simulation-based training · Learning analytics · Big data

✉ Helena J. M. Pennings
h.j.m.pennings@umcutrecht.nl; heleen.pennings@tno.nl

1 Utrecht Center for Research and Development of Health Professions Education, University Medical Center Utrecht, Utrecht, The Netherlands

2 Netherlands Organization for Applied Scientific Research, TNO, Soesterberg, The Netherlands

3 National School of Education, Australian Catholic University, Melbourne, Australia

4 Student Educational Sciences, Radboud University Nijmegen, Nijmegen, The Netherlands

5 Yuverta, Den Bosch, The Netherlands

Published online: 23 May 2025

1 Introduction

Since Bandura introduced the concept of self-efficacy in 1977, many researchers have investigated the role of self-efficacy in different contexts including academic context, specifically in learning. Confidence in the ability to learn and perform specific tasks is important for academic success (e.g., Bandura, 1997; Paschke et al., 2020; Talsma et al., 2019). Bandura defined self-efficacy as "people's beliefs about their capabilities to produce designated levels of performance that exercise influence over events that affect their lives" (1994, p. 71). Self-efficacy is a powerful and widely used predictor of academic achievement and persistence (Choi, 2005; Marsh et al., 2019; Robbins et al., 2004). Although the motivational strength and predictive value of self-efficacy across various domains have been confirmed by a large body of empirical evidence, self-efficacy retains its currency as a concept of interest. Especially considering the new educational tools arising in classrooms and other learning settings, such as Virtual Reality, Serious Gaming, and e-learning, the need for research on the role of self-efficacy continues to exist (Antonopoulos et al., 2024; Bradley et al., 2017). A particular phenomenon of interest is the effect of miscalibrating self-efficacy, or self-efficacy bias, on actual performance (e.g., Paschke et al., 2020; Talsma et al., 2019; Vancouver & Kendall, 2006).

2 General Self-Efficacy and Task-Specific Self-Efficacy

People need to believe that they can produce desired effects with their actions, and people without this belief have little incentive to act (Bandura, 1997). Bandura proposed four sources of self-efficacy: mastery experiences (having a direct experience of mastering knowledge or skill); vicarious experiences (through observing people, who are similar to oneself, achieving success); verbal persuasion (assurance and encouragement from influential people such as parents, teachers, or coaches); and physiological and affective states (positive body's physical and emotional reactions to certain situations). (Bandura, 1986, 1997; Huang et al., 2020; Usher & Pajares, 2009). According to Bandura, it is not the actual ability but an individual's belief about their ability that results in them engaging in specific behavior of executing a specific task. Experiencing success when engaging with any unrelated tasks increases their self-efficacy, while failure lowers self-efficacy for future encounters with similar tasks (Bandura, 1997).

According to Social Cognitive Theory (Bandura, 1986, 1997), self-efficacy beliefs vary across three dimensions: level (the specific difficulty level of a task), strength (the degree of certainty in successfully performing a given level of difficulty), and generality (the extent to which level and strength apply across different tasks and situations). Bandura conceptualized self-efficacy as situation- and task-specific, which some researchers argue has restricted self-efficacy research primarily to its level and strength dimensions (e.g., Chen et al., 2001). Subsequently, the concept of General Self-Efficacy (GSE) has been capturing attention, which is defined as "individuals' perception of their ability to perform across a variety of different situations" (Judge et al., 1998, p. 170). Unlike Bandura's task-specific self-efficacy, GSE is a situation-independent competence belief (Scherbaum et al., 2006). GSE is influenced by all past successes and failures, rather than a task-specific belief. Self-efficacy, specifically the GSE is a relatively stable trait, i.e., once established it becomes harder to change (Sherer et al., 1982).

Bandura emphasized that self-efficacy cannot be considered as a single trait but consists of a differentiated set of beliefs connected with specific tasks and functions (Bandura, 2006), in turn resulting in domain-specific self-efficacy research (Pajares, 1997). While the task and situation specific nature of self-efficacy is underscored (Bandura, 1978), which means that self-efficacy is task-specific and self-efficacy beliefs may vary depending on the context in which someone performs a particular task (Pajares & Schunk, 2005), the role of general self-efficacy is not completely negated. Rather, Bandura agreed that high self-efficacy and mastering a specific task could result in high self-efficacy for other but related tasks, indicating self-efficacy is generalizable and transferable to other contexts (Bandura, 1997), and can be perceived as a trait like concept (Choi, 2005; Marsh et al., 2019). Empirical evidence on the interplay between general and task-specific self-efficacies remains limited. However, a study by Chen et al. (2001) found that their effects may vary in magnitude.

It is widely established that self-efficacy has an influential role to play in relation to human performance, particularly physical (e.g., Feltz et al., 2008) and cognitive (e.g., Richardson et al., 2012) performances. Longstanding studies have reported that self-efficacy beliefs influence the choices people make and the course of action they pursue (Bandura, 1997); the effort they make for engaging in an activity, and the extent to which they persist in an activity while facing obstacles (Bandura, 1997; Pajares, 1996). However, there are studies reporting a negative relationship between self-efficacy and performance in a learning context (e.g., Jernigan, 2004; Vancouver & Kendall, 2006). There is empirical evidence substantiating that the effect of self-efficacy may change over time. For example, some studies have identified the relationship between self-efficacy and performance strengthened over time (e.g., Vancouver et al., 2002) while others have reported the weakening relationship between self-efficacy and performance over time (Mitchell et al., 1994) or no change over time (Lee & Klein, 2002). Different reasons for these conflicting results have been studied extensively, including critical analyses of the self-efficacy measurement scales. Researchers have also recognized the moderation effects of other constructs impacting the predicted relationships between self-efficacy and the outcome variables. A recent study by Horcajo (2022) identified greater effect of self-efficacy on physical and cognitive performance of the participants with higher metacognitive certainty compared to those with lower metacognitive certainty. The study operationalized metacognitive certainty as " as a person's metacognitive assessment that a thought, feeling, belief or attitude is valid, clear in one's mind, or correct" (DeMarree et al., 2020, p. 1239). Schoenherr et al. (2024) argue that looking at self-efficacy (or subjective confidence) neither helps learners to understand the limits of their capacity, nor is a robust measure of learning and actual performance on a task. Furthermore, it needs to be acknowledged that Bandura didn't overlook the potential chances for self-efficacy beliefs being inaccurate in relation to the actual competence of an individual (Bandura, 1986). In such instances of inaccurate self-efficacy beliefs, it is possible that an individual could think either in self-enhancing or in self-debilitating ways. The plausible explanation for negative effects of self-efficacy on performance could be understood in terms of the '*self-efficacy bias*', which is the result of the miscalibration of perceived self-efficacy and actual performance (Boekaerts & Rozendaal, 2010; Talsma et al., 2019, 2020; Vancouver & Kendall, 2006). Bias indicates the direction of judgement error, or in other words, it is just the opposite of accuracy in self-efficacy assessment (the degree to which one's self-efficacy strength matches their actual ability). It means that 'accurate learners' skillfully judge their actual competence and performance while for inaccurate learners, there will be a discrepancy between their judged competence and actual performance.

3 Self-Efficacy (Mis)calibration

A strong sense of self-efficacy does not guarantee high performance rates in adult learners, which could be interpreted as due to the self-efficacy bias or in other words the miscalibration of self-efficacy belief to the actual ability (Schoenherr et al., 2024; Talsma et al., 2019). *Calibration* is defined as "the degree to which judgments of performance accurately reflect actual performance" (Bol et al., 2005, p. 270) or "the relation between confidence and performance or between predicted and actual performance" (Lin & Zabrucky, 1998, p. 347). Calibration is calculated by comparing self-efficacy scores to performance on a task (Klassen, 2002). When self-efficacy judgements and actual performance align this is called accurate *calibration* (Klassen, 2002; Schoenherr et al., 2024). However, calibration of self-efficacy can also be inaccurate (Gonida & Leondari, 2011). Inaccurate calibration or miscalibration of self-efficacy can occur in two forms: (1) *overestimation*, characterized by overconfidence and exaggerated self-efficacy, which is the positive bias and (2) *underestimation*, characterized by uncertainty and underestimation of capabilities, which is the negative bias (Bol et al., 2005; Gonida & Leondari, 2011; Paschke et al., 2020; Talsma, et al., 2019).

While it is important to possess accurate self-efficacy beliefs which is free of bias, it has been reported that a mild overestimation of capabilities (slightly positively biased self-efficacy) can be facilitative and help to motivate students to put in the necessary effort to learn to be consistent in performance, and to be persistent with dealing the problems (Bandura, 1986; Chen, 2003; Gonida & Leondari, 2011; Zimmerman, 2000). However, research on self-efficacy bias and accuracy has shown that not all students are realistic in assessing their self-efficacy or just slightly biased, rather, students often severely overestimate or underestimate their capabilities, which means that calibration accuracy is very low (e.g., Dupeyrat et al., 2011). Considering this tendency to overestimate or underestimate their self-efficacies, it could be interpreted that learners with reported low or high self-efficacies are equally able to succeed in a task when they really want to. Therefore, the goal should be to achieve an accurate calibration of self-efficacy, rather than high self-efficacy at all costs (Stankov & Lee, 2017). However, as mentioned earlier, there is an increased possibility for learners with slightly positively biased self-efficacy to be more successful in completing the task, even if they face any difficulty or setback, while those with low self-efficacy are more likely to give up (e.g., Bandura, 1997; Pajares, 1997).

Miscalibration can have a negative impact on academic performance. For example, *overestimation* may lead to risky behavior (Stankov & Lee, 2017); underestimation of the effort and preparation needed to succeed (Boekaerts & Rozendaal, 2010; Talsma et al., 2019); lower motivation resulting in reduced study efforts(Jernigan, 2004); decreased tendency to seek help (Zvacek et al., 2015); increased carelessness (Vancouver et al., 2002); or, a lack of a crucial realistic basis to approach learning (Moore & Healy, 2008). In terms of mathematical self-efficacy, it has been reported that overestimation results in lower mathematical achievement (Chiu & Klassen, 2010). The study by Ehrlinger et al. (2008) has observed an interesting trend that students performing in the lowest 25% among the cohort tend to overestimate their ability while those who are in the top 25% are more prone to underestimate their ability. There are other studies reporting similar findings, where higher achieving students, although display more accuracy, tend to be underconfident compared to the lower achieving students who are overconfident but less accurate in their judgement of self-efficacy (Miller & Gearaci, 2011). In general, miscalibration is reported to be found in lower performing students, who tend to overestimate their skills, being unskilled and unaware (Bol et al., 2005; Multon et al., 1991; Pajares & Kranzler, 1995; Kruger & Dunning, 1999; Talsma et al., 2019). Overestimation in lower-performing students is more prevalent at task level (e.g., written assignments and exams; Talsma et al., 2019) but is also found at the more general domain level (e.g., course or subject outcome). This contrasts with higher-performing students, whose calibration accuracy is better at task-level than on the general domain level, on which they tend to overestimate their abilities (Talsma et al., 2020). In contrast to overestimation, *Underestimation* may lead to either avoiding engaging in challenging or new tasks, over-studying, assigning unjustified resources, adopting unhelpful suggestions, or disengagement from learning, because of uncertainty and self-doubt (Boekaerts & Rozendaal, 2010; Talsma et al., 2019; Usher, 2016), all of which prevent mastering or developing new skills.

Researchers call for further research on miscalibration of self-efficacy in adult learners, and in authentic learning environments (e.g., Talsma et al., 2019). To train in near authentic environments in the military, often (high fidelity) Simulator Based Training (SBT) is used. Considering that t there is not much research reported on the (mis)calibration of self-efficacy in a military training environment, the findings from this study contribute to addressing the gap in self-efficacy research.

4 Self-Efficacy and Simulator-Based Training

Simulator-Based Training is a technology enhanced learning environment that is often used in military training or other safety critical domains, such as medicine, shipping, aviation, and the chemical industry (e.g., Dahlstrom et al., 2009; Nazir et al., 2015). Simulators offer a learning environment in which complex and dangerous tasks can be trained in a safe, controlled, and realistic environment (Bell et al., 2008; Marcano et al., 2019). It is a relatively economic option to practice the relevant skills in high-risk situations (Dahlstrom et al., 2009), and it enhances students' confidence and competence (Dunn et al., 2014). SBT is therefore more motivating than traditional learning environments, especially concerning self-efficacy (Gegenfurtner et al., 2014). Also, SBT has pedagogical advantages, as the training can be adjusted to the level of the students' current competence (Maran & Glavin, 2003). However, maximizing the effectiveness and efficiency of SBT is crucial (Dahlstrom et al., 2009) to create maximum benefit and to avoid any potential economic burden. Hence, a comprehensive understanding of the relevant factors that affect the learning processes in SBT is very important. With the present study we aim to deepen our understanding of how generalized self-efficacy, task-specific self-efficacy and self-efficacy bias may affect performance in a high-fidelity military simulator. Accordingly, we extracted data on learners' performance in the simulator, including the number of errors (categorized by severity) and the final score, from the simulator log files.

Self-efficacy is often used a measure of simulator training effectiveness. Yet, the evidence regarding the relationship between SBT and self-efficacy remains inconclusive. While there are several studies reporting a positive association between performance and self-efficacy in SBT environments (Backlund et al., 2008; Bell et al., 2008; Davis et al., 2000; Dunn et al., 2014; Maenhout et al., 2021; Orvis, et al., 2009), there are studies reporting no association between self-efficacy and performance (Lee et al., 2020). Similarly, experimental studies have yielded mixed results. For instance, some studies have reported higher self-efficacy in SBT groups compared to control groups (Hsu et al., 2015; Sitzman, 2011), while others have found no significant differences in self-efficacy between

SBT and non-SBT groups (Bragard et al., 2019; Luctkar-Flude et al., 2012; Roh et al., 2013). Stellflug and Lowe (2018) suggested that self-efficacy might not be a good indicator of (future) performance, especially since simulation training and research has evolved beyond self-assessment. For example, simulator log data could be used as a more robust measure of performance.

The large amount of data stored in simulators contains valuable information that can be used to improve and personalize the learning experience (Kizilcec et al., 2013). These data may also provide more objective performance indicators that add to the existing knowledge about the miscalibration of self-efficacy. This information is more objective than instructor observations or the trainees' own observations, because the actual actions of the trainees are logged automatically. Thus, whether a trainee has performed a certain task correctly or incorrectly depends on certain preset norms to which the logged trainees' actions are compared.

Most of the research exploring the relationship between self-efficacy and SBT have been conducted in health professions education (e.g., Bragard et al., 2019; Dunn et al., 2014; Fadale et al., 2014; Hsu et al., 2015; Maenhout et al., 2021; Morfoot & Stanley, 2018; Roh et al., 2013; Sun et al., 2017). There is hardly any research that focused on studying self-efficacy in simulation based military training. We were able to find three studies: two of which used serious games as SBT (Orvis et al., 2009; Whitney et al., 2012), and the third study compared SBT with another training environment (Barnes et al., 2016).

Orvis et al. (2009) studied the performance and self-efficacy of military cadets in a serious gaming environment. A positive association between pre-training gaming self-efficacy and post-training gaming performance was found. However, the results of a multiple regression analysis endorsed gaming experience as the significant predictor of post-training gaming performance. Whitney et al. (2012) studied general and task-specific self-efficacy for countering the threats of Improvised Explosive Devices (IEDs or Counter-IED) of military personnel in a serious gaming environment (VBS2) and traditional drill environment. Higher task-specific self-efficacy was reported in the traditional drill environment compared to the gaming environment. Participants described the game as very difficult, which resulted in frustration and probably lower self-efficacy. Barnes et al. (2016) examined the surgical skills training of military medical personnel, comparing training in a live tissue environment to a simulated environment. They found that simulation had a positive effect on self-efficacy, but only for novices. Shortcomings were found in both training types, therefore, Barnes and colleagues concluded that a combination of both may be most beneficial for novices.

Our extensive literature review, revealed a lack of research published exploring the relationship between self-efficacy and high-fidelity simulators or studies on miscalibration of self-efficacy in SBT training. This context underscores the significance of the current study.

5 Research Questions and Hypotheses

The reported studies on self-efficacy and SBT either explored the change in self-efficacy after using SBT or compared self-efficacy across different types of training. However, it has been reported that a strong sense of self-efficacy does not guarantee high performance (Talsma et al., 2019). For instance, *overestimating* ability could result in carelessness and risky behavior (Stankov & Lee, 2017) or a lack of a crucial realistic basis (Moore & Healy, 2008) and *underestimating* ability could result in avoiding engaging in challenging or new tasks, adopting unhelpful suggestions, or disengagement, all of which prevent mastering or developing new skills (Boekaerts & Rozendaal, 2010; Talsma et al., 2019; Usher, 2016). Both are very dangerous outcomes within a military context. Therefore, instead of aiming to increasing self-efficacy, the goal should be to achieve an accurate calibration (Stankov & Lee, 2017). Considering the gap in the literature on self-efficacy and (mis-)calibration in military training settings, the current study is focused on the (mis-)calibration of self-efficacy in a near authentic (high fidelity) SBT environment analyzing calibration accuracy of military trainees.

This study can be considered as an initial attempt to move a few steps forward exploring the influence of general self-efficacy, task-specific self-efficacy, and self-efficacy bias in a high-fidelity military training simulator called the Combat Vehicle 90 (CV90) simulator.

The following research questions are investigated in this study:

(i) How are general self-efficacy and task-specific self-efficacy related to overall performance in the simulator?
(ii) Are trainees able to accurately calibrate their general and task-specific self-efficacy? If not, are trainees overestimating or underestimating their self-efficacy?
(iii) Do overestimating, accurately estimating, and underestimating trainees significantly differ in the number of *high,medium,*and *low severity* errors they make?

Based on the studies concluding a positive relationship between self-efficacy and performance (Bandura, 1997; Dixson et al., 2016; Robbins et al., 2004), we hypothesized positive relationships between both general and task-specific self-efficacy, with overall performance (i.e., *end score*). However, given that self-efficacy is primarily task-specific (Bandura, 1998; Marsh et al., 2019), it is hypothesized that the correlation between task-specific self-efficacy and performance is stronger than the correlation between general self-efficacy and performance. Also, we expect that general self-efficacy and task-specific self-efficacy are positively, but not perfectly correlated (Marsh et al., 2019). Performance on specific tasks is measured by the number of errors made, so it is hypothesized that both general and task-specific self-efficacy are negatively related to the performance indicators. Based on the studies by Talsma et al., (2019, 2020), it is assumed that about two-third of trainees miscalibrate their self-efficacy, with the majority overestimating their general self-efficacy, but underestimating the task-specific self-efficacy. Furthermore, it is hypothesized that miscalibration is especially found in lower performing trainees, who tend to overestimate (i.e., both in general and task-specific self-efficacy). Therefore, it is expected that the overestimating trainees make more errors of all severity types than their underestimating counterparts who are more accurate in estimation.

6 Method

6.1 CV90 Gunner Training

The CV90 is an armored infantry combat vehicle. The crew consists of three people: the driver, the gunner, and the commander. The CV90 is equipped with several weapons (e.g., canon, coaxial machine gun, several types of grenades). The gunner in a CV90 has the responsibility to protect the CV90 and the infantrymen it is transporting. The gunner operates all weapons according to the appropriate procedures, that are being learned in a high-fidelity simulator which is a realistic copy of the actual vehicle. This training consists of 24 lessons with Computer Assisted Instruction (CAI) that guides trainees through the lessons. Every lesson consists of multiple short lesson steps that are either instructional or practical. In the practical lesson steps trainees execute short scenarios. The two possible areas for trainees making errors in practical lessons are either wrongly executing a task or failing to eliminate the target within the time-limit.

Whether a practical lesson step is completed successfully or not depends on the severity of the errors made during the lesson step. A *high severity* error results in immediate abortion and repetition of the lesson step; after a *medium severity* error the trainee is allowed to finish the lesson step but has to repeat the lesson step afterwards; *low severity* errors consist of redundant, but usually harmless actions, and have no consequences (trainees receive no feedback about these errors, neither through the CAI nor from the instructors). In regular lessons, lesson steps can be repeated multiple times, until the lesson step is completed with no *high* or *medium severity* errors being made. In test lessons, lesson steps cannot be repeated. After each lesson an end score is calculated that represents overall performance during the lesson.

6.2 Participants

Data were collected from trainees ($N=66$) of the Dutch Defense Organization, who were training to become gunners in the CV90 armored vehicle. Majority of the participants were male (98.5%). Trainees had different educational backgrounds: secondary education ($n=6$), vocational education ($n=48$), academic bachelor ($n=10$), and academic master ($n=2$). Most trainees were in their late teens/early twenties, but their exact age is not known to the researchers. Data were collected at three points: before, during, and after the training of four groups of trainees during 2018–2019 (May 2018, September 2018, February 2019, and September 2019).

6.3 Instruments

6.3.1 General Self-Efficacy

The eight item self-efficacy subscale of the Motivated Strategies for Learning Questionnaire (MSLQ; Pintrich et al., 1993) is a well-validated and widely used questionnaire in educational research (Hilpert et al., 2013), as well as military education (e.g., Nurnindyah, et al., 2023). Although the MSLQ could have been administered in its entirety, we preferred the option to choose the self-efficacy subscale to measure general self-efficacy. As described in the MSLQ manual, the eight items were rated on a 7-point Likert type scale (*1* = not at all true and *7* = very true). Some terminologies in the items were altered to fit with the terms used in the military training context (e.g., training instead of class, and instructor instead of teacher). These alterations did not change the meaning of the items. Example items are *"I expect to do well in this training"* and *"I am certain I can master the skills being taught in this training"*. The internal consistency of this scale with this sample was good (reliability coefficient Cronbach's $\alpha = 0.83$).

6.3.2 Task-Specific Self-Efficacy

Task-specific self-efficacy was operationalized as trainees' perceived ability to learn and perform specific subtasks to be an effective gunner in the CV-90. Examples of these subtasks include calibrating the setup of the system, eliminating static and moving targets, eliminating (static and moving) targets from a static or moving vehicle, choosing the correct ammunition to eliminate a target, and, eliminating target with poor visibility (e.g., in the dark). Based on these tasks, together with the instructors, we developed a new instrument to measure trainees' self-efficacy in performing these tasks.

This resulted in a new ten item rating scale, the 'Perceived Gunner Ability Questionnaire' and measured perceptions of trainees about their gunner ability as one construct. Items were rated on a 7-point Likert scale (*1* = I strongly disagree and *7* = I strongly agree). The items were related to the tasks that were being trained; Example items are *"I know where to find the different instruments in the vehicle*" and *"I am capable of eliminating a target from a moving vehicle"*. Task-specific self-efficacy was measured three times: Prior to the training (i.e., pre-training), half-way during the training (i.e., mid-training) and at the end of the training (i.e., post-training). Internal consistency for this scale in this sample was good, with reliability coefficient Cronbach's alphas of 0.94, 0.88, and 0.96 for the three timepoints, respectively.

6.3.3 Overall Performance

Overall performance was measured using data retrieved from the simulator logs: We used the average number of *high*, *medium*, and *low severity* errors made by trainees across lesson training and the average *End score* reflecting the overall performance of trainees across the entire training in the simulator. The end score is calculated in the simulator as the percentage of unique lesson steps in which no *high severity* or *medium severity* errors are made. The end score was also retrieved from the simulator log files.

6.4 Procedure

This study was approved by the ethical review board of The Netherlands Organization for Applied Scientific Research, case number 2018–019. Prior to the collection of the data, trainees received a verbal explanation about the purpose and content of the study along with written information about the study. Participation in the study was purely voluntary, and trainees could drop out of the study at any time, without any consequences. All trainees actively consented to participate, and none dropped out. The trainees received a personal participant number before the start of the data collection; no personal information about the trainees was known to the researchers.

Prior to the start of the first lesson in the simulator, data on general and task-specific self-efficacy were collected using pen and paper questionnaires. During the training, all actions made by the trainees were logged by the simulator and extracted after the course ended. Mid-training and post-training measurements of task-specific self-efficacy were collected right before the test lessons, to avoid any influence of the test lessons results.

6.5 Data Handling and Analysis

All analyses were performed in IBM SPSS statistics (Version 25). To answer the research questions, the data on general self-efficacy, task-specific self-efficacy, and end score were re-coded into new variables. First, the variable *end score* was re-coded into 7 categories using Visual Binning. The scores on general and the three task-specific self-efficacy variables, respectively, were then subtracted from the new *end score* variable, resulting in the four continuous variables ranging from -7 to 7, which represented the calibration accuracy: *accuracy of general self-efficacy* and *accuracy of task-specific self-efficacy* (separate variables for pre-training, mid-training, and post-training). Trainees were then sorted into three groups, representing 'overestimating' (i.e., trainees scoring ≤ -2), 'accurately estimating' (i.e., trainees scoring between -2 and 2), and 'underestimating' (i.e., scoring ≥ 2). This was done separately for all four variables, resulting in *calibration level of general self-efficacy* and *pre-training calibration level of task-specific self-efficacy, mid-training calibration level of task-specific self-efficacy*, and *post-training calibration level of task-specific self-efficacy.*

Three Chi-square tests were used to test whether trainees' general self-efficacy and task-specific self-efficacy fell in the same category, separately for pre-training, mid-training, and post-training.

To test whether task-specific self-efficacy changes over time, and whether this differs for the three pre-training calibration level groups a repeated measures ANOVA was performed. Pre-training self-efficacy, mid-training self-efficacy and post-training self-efficacy were included as repeated measures of task-specific self-efficacy. The three levels of calibration identified for pre-training self-efficacy calibration were included as the grouping variable. An additional Kruskal–Wallis H test was used to test the significant differences in pre-training task-specific self-efficacy.

Before testing whether trainees in the different calibration groups differed in the number of mistakes they made, we first checked the assumptions for performing one-way ANOVAs. For general self-efficacy, the assumptions of normality and heterogeneity were violated in several variables. We decided to perform non-parametric tests to test the differences in number of errors and end scores between the calibration groups. As there were three calibration groups in the pretest, the differences in number of errors between the calibration groups of general self-efficacy and pre-training task-specific self-efficacy were tested with the Kruskal–Wallis *H* test. As there were only two calibration groups identified in the mid-training and post-training, these differences were tested using the Mann–Whitney *U* test for two independent groups. In all analyses, the dependent variables were *high severity errors*, *medium severity errors*, *low severity errors,* and *end score*.

6.5.1 Missing Data

Due to a logistical error, the second group of trainees ($n = 15$) did not complete the Perceived Gunner Ability Questionnaire prior to the start of the training. The data were therefore considered missing and excluded from the analyses concerning pre-training task-specific self-efficacy. Also, some trainees did not complete the mid-training ($n = 3$) and post-training self-efficacy measure ($n = 5$), for reasons that remain unclear. These participants were only excluded for the analyses of mid-training and post-training task-specific self-efficacy.

7 Results

7.1 Correlations

The correlations between the simulator log data (severity and type of errors, and end scores) and the questionnaire data (general and task-specific self-efficacy) are presented in Table 1. Weak to moderate positive associations were found between general self-efficacy and the three measures of task-specific self-efficacy. These correlations gradually decreased over time. Pre-training task-specific self-efficacy was not correlated to mid-training and post-training task-specific self-efficacy, but the latter two were strongly correlated with each other. General self-efficacy and the mid-training task-specific self-efficacy were significantly correlated to the *end score*. Furthermore, only a few significant, but weak correlations were found between the measures of self-efficacy and task errors.

7.2 Calibration of General and Task-Specific Self-Efficacy

Figure 1 shows the percentage of trainees in each of the calibration groups per time point. For calibration of general self-efficacy, the results showed that 54.5% of the trainees were able to accurately calibrate their general self-efficacy, and the remaining 45.5% of the trainees miscalibrated their self-efficacy of which 7.6% were underestimating trainees and 37.9% were overestimating trainees. A similar pattern was found for calibration of pre-training task-specific self-efficacy: 51% of the trainees had an accurate calibration and 49% of the trainees miscalibrated their self-efficacy with 9.8% underestimating trainees, and 39.2% overestimating trainees.

For mid-training only two groups were identified. Several trainees adjusted their calibration and fell into another category compared to their pre-training calibration data. For mid-training calibration, 47.6% of the trainees were overestimating and 52.4% had accurate calibration. For post-training calibration only one participant was underestimating. After testing the effect of removing this participant or adding this participant to the accurate-calibration group, we decided to keep this participant in the analysis as an accurately

Table 1 Correlations between simulator log data, end score, questionnaire data, accuracy of general self-efficacy, and accuracy of task-specific self-efficacy

		1	2	3	4	5	6	7	8
1	High severity errors	–							
2	Medium severity errors	0.66***	–						
3	Low severity errors	0.52***	0.40**	–					
4	End score	−0.85***	−0.81***	−0.41**	–				
5	General self-efficacy	−0.25*	−0.25*	−0.14	0.32**	–			
6	Task-specific self-efficacy pre-training	−0.14	−0.14	−0.02	0.14	0.44**	–		
7	Task-specific self-efficacy mid-training	−0.27*	−0.24	−0.17	0.28*	0.36**	0.18	–	
8	Task-specific self-efficacy post-training	−0.25	−0.23	−0.12	0.24	0.27*	0.25	0.59***	–

*$p<0.05$; **$p<0.01$; ***$p<0.001$

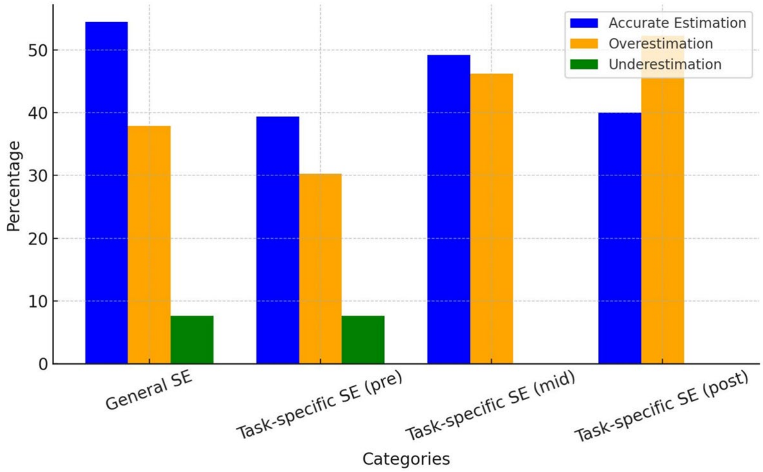

Fig. 1 Bar charts of percentage of participant in the calibration groups per measurement

calibrating trainee (i.e., no differences were found). This resulted in 55.7% of overestimating and 44.3% of accurately calibrating trainees.

The results of a Chi-square test showed that the distributions of *calibration level of general self-efficacy* and *calibration level of pre-training task-specific self-efficacy* were strongly related (Cramer's V = 0.67, $p < 0.001$), but significantly different $\chi(4) = 45.75$, $p < 0.001$.[1] Thus, having a certain calibration level for general self-efficacy did not necessarily result in the same calibration level for task-specific self-efficacy. Similar patterns were found for pre-training with mid-training (Cramer's V = 0.66, $p < 0.001$; $\chi(2) = 20.68$, $p < 0.001$) and mid-training with post-training (Cramer's V = 0.88, $p < 0.001$; $\chi(2) = 46.88$, $p < 0.001$) self-efficacy calibration levels; some overestimating trainees became accurately calibrating, some accurately calibrating trainees became overestimating, and all underestimating trainees became accurately calibrating trainees.

7.3 Change in Self-Efficacy Over Time

The results of the repeated measure ANOVA showed a significant linear effect and a significant quadratic effect of task-specific self-efficacy over time (F_{linear} (1, 43) = 48.33, $p < 0.001$, partial $\eta_p^2 = 0.53$ and $F_{quadratic}$ (1, 43) = 19.46, $p < 0.001$, partial $\eta_p^2 = 0.31$). These effects were not moderated by pre-training calibration level. Between subject effects were only found for the intercept (F (1, 43) = 2829.88, $p < 0.001$, partial $\eta_p^2 = 0.99$). An additional one-way ANOVA revealed that the *pre-training task-specific self-efficacy* of underestimating trainees ($M = 3.99$) was significantly lower than both the accurate ($M = 4.89$) and overestimating trainees ($M = 5.31$).

[1] Note that this test could only be performed on the 51 trainees who completed both questionnaires for general and pre-training task-specific self-efficacy.

7.4 (Mis)calibration and Performance

Table 2 provides the Means and Standard Deviations for the simulator log data (*severity errors* and the *end scores*) and the questionnaire data (general self-efficacy and pre-, mid-, and post-training task-specific self-efficacy) for the entire group of participants (i.e., Overall) and separately for the calibration groups per measurement. In Fig. 2 boxplots are presented corresponding to the distributions of the number of errors each of the calibration groups make at the three time-points. The mean values indicate that, for all severity errors, the number of errors was higher for overestimating trainees compared to accurate and underestimating trainees.

The differences in number of errors between the calibration groups of general self-efficacy and pre-training task-specific self-efficacy were tested with the Kruskal–Wallis H test (Table 3). For general l self-efficacy, all error types showed a significant difference between the overestimating and accurate trainees but not between the accurate and underestimating trainees. For *pre-training task-specific self-efficacy*, *overestimating* trainees made significantly more *high severity* and *medium severity errors* than accurate and underestimating trainees. For *End score* also the difference between underestimating and accurate were significant, both for general self-efficacy and for task-specific self-efficacy pre-training. Thus,the underestimating trainees were those with the highest *End scores*. The effect sizes were medium to large.

As there are only two calibration groups identified for mid-training and post-training, we conducted the Mann–Whitney *U* test for two independent groups to test the differences in error types and end score for these groups (Table 4). For *mid-training task-specific self-efficacy* and *post-training task-specific* self-efficacy the results were similar to the results for *general self-efficacy*. Overestimating trainees made significantly more errors than accurately estimating trainees on all severity levels. Also, trainees in the accurate group received significantly higher end scores compared to those in the overestimating group.

8 Discussion

With the present study we explored the degree of self-efficacy bias that was present in military trainees who followed SBT in a high-fidelity training simulator. We studied whether trainees were able to accurately calibrate their self-efficacy to their performance, and, if not, whether they overestimated or underestimated themselves, and how this was related to the number of high, medium, and low severity errors they made.

8.1 RQ1: Correlations Among General Self-Efficacy, Task-Specific Self-Efficacy, and Overall Performance

We first studied the correlations among general self-efficacy, three measurements of task-specific self-efficacy, and overall performance. We assumed positive relationships among the four measurements of self-efficacy and overall performance, and that the correlations between task-specific self-efficacy and performance would be stronger than general self-efficacy and performance.

As expected, general self-efficacy was positively correlated to pre-, mid-, and post-training task-specific self-efficacies. However, pre-training task-specific self-efficacy was

Table 2 Means and standard deviations averaged over all participants, and per subgroup

		General self-efficacy (N = 66)			Pre-training			Mid-training		Post-training	
					Task-specific self-efficacy (N = 51)			Task-specific self-efficacy (N = 63)		Task-specific self-efficacy (N = 61)	
	Overall	OE (N = 25)	AE (N = 36)	UE (N = 5)	OE (N = 20)	AE (N = 26)	UE (5)	OE (N = 30)	AE (N = 33)	OE (N = 34)	AE (N = 27)
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
High severity	5.10 (1.57)	6.24 (1.57)	4.46 (1.02)	3.33 (0.67)	6.01 (1.18)	4.31 (0.10)	3.76 (1.27)	5.90 (1.53)	4.17 (1.07)	5.83 (1.50)	4.02 (1.00)
Medium severity	7.92 (2.40)	9.65 (2.42)	7.11 (1.62)	5.07 (0.86)	9.30 (1.70)	7.25 (1.85)	6.29 (0.83)	9.14 (2.55)	6.75 (1.65)	8.96 (2.50)	6.61 (1.66)
Low severity	7.14 (3.09)	8.60 (3.41)	6.43 (2.59)	4.91 (1.77)	8.09 (3.36)	6.52 (2.94)	6.42 (2.69)	7.88 (3.26)	6.21 (2.60)	7.62 (3.15)	6.17 (2.84)
End score	69.08 (6.12)	63.26 (5.23)	72.08 (2.96)	76.64 (1.89)	64.12 (3.32)	72.11 (3.77)	75.03 (3.40)	64.61 (5.41)	73.40 (3.03)	65.31 (5.45)	73.79 (3.09)
General self-efficacy	5.02 (0.61)	4.93 (0.63)	5.12 (0.60)	4.64 (0.18)	4.98 (0.58)	5.06 (0.63)	4.64 (0.24)	4.93 (0.67)	5.12 (0.54)	4.99 (0.67)	5.07 (0.55)
Task-specific self-efficacy pre-training	4.89 (1.04)	5.00 (0.90)	4.80 (1.22)	4.87 (0.68)	5.29 (0.82)	4.80 (1.07)	3.77 (0.90)	4.86 (1.02)	5.10 (0.91)	4.94 (1.07)	5.04 (0.88)
Task-specific self-efficacy mid-training	5.89 (0.60)	5.92 (0.62)	5.85 (0.60)	5.98 (0.52)	5.89 (0.67)	5.93 (0.49)	5.93 (0.59)	5.93 (0.59)	5.84 (0.61)	5.88 (0.58)	5.93 (0.61)
Task-specific self-efficacy post-training	6.12 (0.63)	6.14 (0.61)	6.09 (0.69)	6.18 (0.41)	6.13 (0.82)	6.11 (0.42)	6.00 (0.78)	6.19 (0.58)	6.04 (0.68)	6.23 (0.58)	5.97 (0.68)

OE, Overestimating group; AE, Accurate-efficacious group; UE, Underestimating group; NA, Not Applicable

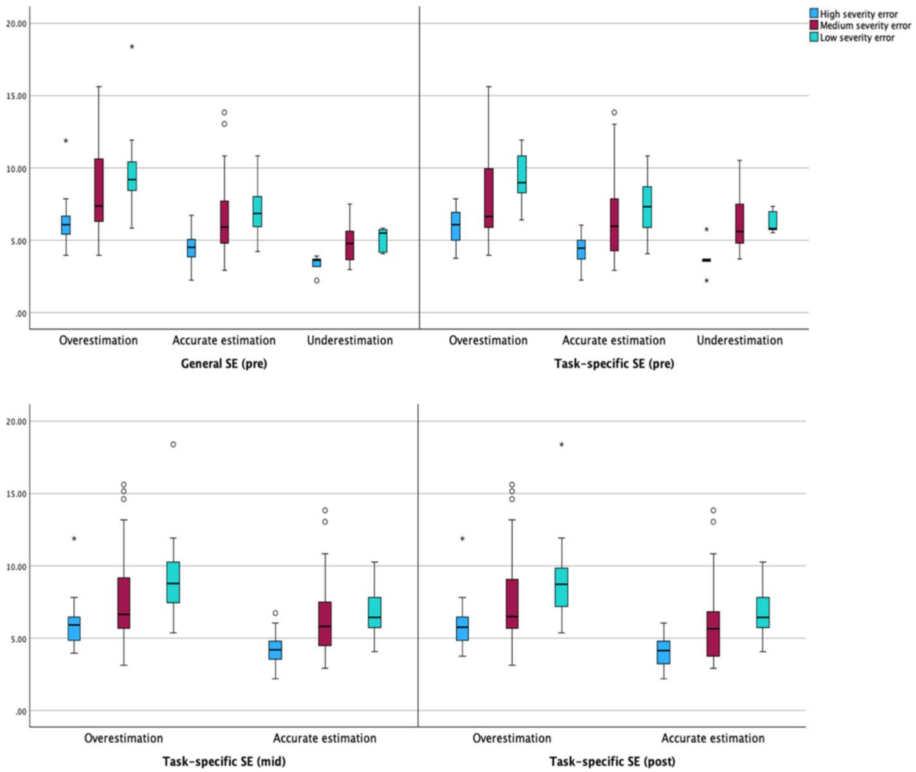

Fig. 2 Boxplots corresponding to the number of different errors participants in the different calibration groups made, separate for measurement moment

not related to both mid-training and post-training task-specific self-efficacy, whereas the latter two were strongly correlated. According to Schunk (1996), when skills are already acquired, self-efficacy is the same as perceived ability. However, when skills are yet to be acquired, self-efficacy refers to the belief one can learn a certain skill and is thus a prediction of the expected outcome. The pre-training measurement of task-specific self-efficacy in our study represents such prediction, and for trainees who had never been exposed to this type of SBT, predicting performance in this simulator may have been difficult. This was reflected in the significant between-subject differences found for pre-training task-specific self-efficacies.

The results showed a weak, but significant relationship between overall-performance (i.e., end score) and general self-efficacy, as well as mid-training task-specific self-efficacy. In contrast to our expectation, pre-training and post-training task-specific self-efficacies were not related to the overall performance. The mid-training and post-training measurements reflected the trainees' perceived ability, as they already knew more about their performance in the simulator. This could, therefore, explain the absence of the expected correlations with mid-training and post-training task-specific self-efficacy and overall-performance.

The lack of correlation between post-training task-specific self-efficacy and overall-performance was surprising but might be explained in terms of the content of the training. In

Table 3 Results of the Kruskal–Wallis tests including multiple comparisons

Errors	General self-efficacy (N = 66)							Task-specific self-efficacy pre-training (N = 51)						
	Kruskal–Wallis H test	UE vs A		UE vs. OE		A vs. OE		Kruskal–Wallis H test	UE vs A		UE vs. OE		A vs. OE	
	H^2 (df = 2)	*z*	*r*	*z*	*r*	*z*	*r*	H2 (df = 2)	*z*	*r*	*z*	*r*	*z*	*r*
High severity	27.11***	1.96	0.24	4.30***	0.53	4.51***	0.56	20.78***	0.87	0.12	3.28**	0.46	4.09***	0.57
Medium severity	28.07***	2.19	0.27	4.37***	0.54	4.21***	0.52	15.89***	1.26	0.18	3.20**	0.45	3.31**	0.46
Low severity	11.82***	1.31	0.16	2.77*	0.34	2.81*	0.35	3.74	NA	NA	NA	NA	NA	NA
End score	48.41***	−6.01***	−0.74	−5.21***	−0.64	−2.07*	−0.25	31.75***	−5.06***	−0.71	−4.03***	−0.56	−1.04	−0.15

$^*p < .05$. $^{**}p < .01$. $^{***}p < .001$. UE vs. A, underestimating vs accurate; UE vs. OE, underestimating vs overestimating; A vs. OE, accurate vs overestimating; NA, Not Applicable

Table 4 Results of the Mann–Whitney *U* tests

Errors	Mid-training Task-specific self-efficacy (N = 63)						Post-training Task-specific self-efficacy (N = 61)					
	Overestimating group	Accurate group	Mann–Whitney Us test				Overestimating group	Accurate group	Mann–Whitney U test			
	Median	Median	*U*	*p*	*z*	*r*	Median	Median	*U*	*p*	*z*	*r*
High severity	5.89	4.18	146.50	0.000	−4.80	−0.60	5.77	4.18	117.50	0.000	−4.96	−0.63
Medium severity	8.79	6.43	201.50	0.000	−4.04	−0.51	8.74	6.30	187.50	0.000	−3.94	−0.50
Low severity	6.66	5.83	322.00	0.017	−2.38	−0.30	6.49	5.75	309.50	0.030	−2.17	−0.28
End score	64.53	73.39	969.00	0.000	6.52	0.82	65.80	73.43	890.00	0.000	6.26	0s.80

*$p < .05$. **$p < .01$. ***$p < .001$. NA, Not Applicable

the first half of the training, the lessons focused solely on learning the basic procedures. In the second half, the same procedures were practiced, but trainees sometimes needed to switch to manual mode to perform the tasks, as these lessons involved emergency procedures. Handling the manual settings may result in errors, which trainees could probably quickly resolve due to practice in the first half of the training. However, the *end score* is calculated as the percentage of unique lesson steps without any *high severity* or *medium severity* errors, regardless of the number of errors made. This may cause an imperfect reflection of the actual performance of trainees (see limitation section for our ideas about the end score that is calculated in the simulator).

Furthermore, we expected that both general and task-specific self-efficacy would be negatively related to the task performance indicators (number of errors). Although several significant correlations were found, overall, it could be concluded that the general self-efficacy and the three measurements of task-specific self-efficacy were poor predictors for the performance on specific gunner tasks during this training. Concerning general self-efficacy, this is in line with findings reported by Talsma et al., (2019, 2020). Concerning task-specific self-efficacy, it may have been better to use separate indices for the separate tasks and combine these with task-specific errors made in the simulator, instead of one overall score for task-specific self-efficacy, as we did in the current study.

8.2 RQ2: Calibration of General and Task-Specific Self-Efficacies

The second research question focused on the calibration accuracy of general and task-specific self-efficacies. We expected that some trainees would miscalibrate their self-efficacy to performance, and that the majority of miscalibrating trainees would be overestimating their general self-efficacy, and underestimating their task-specific self-efficacy.

The results showed that approximately half of the trainees were able to accurately calibrate both their general and task-specific self-efficacies. In general, miscalibrating trainees tended to be overestimating rather than underestimating their performance for both general and task-specific self-efficacies. Throughout the training, calibration of self-efficacy changed and three patterns of change were identified: (1) underestimating trainees adjusted their task-specific self-efficacy in such a way that they became accurate-efficacious, (2) some pre-training overestimating trainees adjusted their task-specific self-efficacy to match their performance, and (3) others adjusted their task-specific self-efficacy and became overestimating.

Although the identified trends in general self-efficacy were in line with our expectations, it was not the case for task-specific self-efficacy. This could possibly be explained by the differences in the highly procedural gunner skills that we investigated compared to the less procedural literacy skills studied in previous research (e.g., Talsma et al., 2019). It is possible that novices struggled to grasp the highly procedural gunner skills at an optimal performance level.Also, the focus on negative (i.e., number of errors) as opposed to the positive performance indicators is likely to result in completely different perceptions of performance in our study.

The group composition of the participants also needs to be taken into consideration. In our study, 98.5% of the participants were male, Previous research has shown that male students are significantly more likely to overestimate their abilities than females (Bench et al., 2015; Kallia & Sentance, 2018), while females tend to underestimate their ability (Kallija & Sentence, 2018; Usher et al., 2019; Tzamaras et al., 2024). Therefore, it is not that surprising that trainees in our sample were overestimating their performance.

Also, it is known that males tend to lend their self-efficacy beliefs through vicarious experiences, observing, and comparing themselves with peers (Huang et al., 2020; Usher & Pajares, 2006, 2009). As these trainees received instructor feedback and knew about their peers' end scores and reflections after each lesson; it is likely that the perception of their own performance was affected. Therefore, it is not surprising that the predominantly male group of trainees in our study was likely to overestimate task-specific self-efficacy. We strongly recommend further research to explore this pattern of observation.

8.3 RQ3: (Mis)calibration and Performance

Third, we studied whether the *overestimating, accurate,* and *underestimating trainees* significantly differ in the number of *high severity*, *medium severity*, and *low severity* errors they make and their *end score* they receive. Based on the theory that states overestimation often results in carelessness and reckless behavior (e.g., Stankov & Lee, 2017) and existing empirical evidence, (e.g., Talsma et al., 2019, 2020), we expected that overestimating trainees would make more errors than the accurate and underestimating trainees, in all three error categories (i.e., high, medium, and low severity). The results indeed showed that for both general self-efficacy and the three task-specific self-efficacy measurements, overestimating trainees made more errors than accurate and underestimating trainees. This could indicate that overestimating trainees indeed engage in more carelessness (Vancouver et al., 2002) or risky behavior (Stankov & Lee, 2017). It could also indicate that these trainees underestimate the effort that is needed to succeed (Boekaerts & Rozendaal, 2010; Talsma et al., 2019). Their motivation to improve performance and learning could be lower, because they already think they are doing sufficiently well (Jernigan, 2004). This is idea is also strengthened by the finding that the number over overestimating trainees appears to increase over the course of the three measurements.

The results suggest that overestimating trainees are not the ones who perform well on the task. Instead of scoring high and perceiving their performance as even higher, these trainees score low yet still hold a false perception that their performance levels are high. This result is in line with earlier research (Bol et al., 2005; Multon et al., 1991; Pajares & Kranzler, 1995; Kruger & Dunning, 1999; Talsma et al., 2019) and is especially important for instructors, as these trainees should be given further attention, helping them to improve their performance and to learn to maintain realistic perceptions about their performance, rather than being unconsciously incompetent. Nevertheless, we would like to acknowledge the impact of gender (predominantly males) and context (military training) while interpreting these trends of overestimation in the data. Military trainees want to show that they are tough and can do anything. In the military, weakness or failure is not an option as any mistake in operations and deployments can have life-or-death consequences. However, embracing mistakes as learning opportunities allows military trainees to transform initial weaknesses into strengths through self-regulation and learning.

Furthermore, it is important to consider the nature and process of providing feedback enhancing self-efficacy and monitoring learning (Jahannes & Haasse, 2022). The trainees in the present study received feedback in the simulator, in terms of the most critical errors they made during a lesson step. Later, during the *After Action Review*, trainees received short feedback on their performance, again only in terms of the most critical errors they made, but not on what they did well. When trainees were asked about their performance in the simulator, they often referred to frustrations with the system. Often low performance was attributed to simulator failure or mistakes in scenarios. Instead, if there were opportunities for guided reflections on what they did well and how could they further improve their own actions to achieve better results, it would have benefitted them with their self-efficacy calibration.

9 Limitations and Suggestions for Future Research

We would like to acknowledge the limitations of this study and reflect on the possibilities for future research. The population in this study was predominantly male (98.5%). This means that the results of our study are not generalizable to female trainees. However, when focusing on other military populations (i.e., both from the Dutch Defense Organization or from other countries) our results may be generalizable, since the military population is still predominantly male (Ministerie van Defensie, 2021). However, in future research, gender balance in sample needs to be considered for more generalizable results. It would also provide a potential opportunity to investigate the effect of underestimation which has also been identified as having a negative impact on performance (Boekaerts & Rozendaal, 2010; Usher, 2016).

Another limitation is associated with the calculation of the *end score* in the simulator and the related interpretation of the *general performance*. The *end score* is the percentage of unique lesson steps in which no *high severity* or *medium severity errors* were made. This means that the actual number of errors or the number of lesson step repeats are not incorporated in the *end score*. For future research and for practice, it is advised to create an alternative end score, in which the number of all errors and lesson step repeats are used, so that the *end score* is a better reflection of *overall performance*.

Furthermore, it would have been interesting to have more demographic information about the trainees, such as age and rank. Future research could focus on investigating the development of self-efficacy considering the personal characteristics that may influence self-efficacy and self-efficacy calibration. For example, the effect of the rank for which the trainees are training (i.e., officer, under-officer, or trooper), educational level and age of trainees could be further investigated. The influence of other factors such as self-esteem and reflection skills on self-efficacy and calibration of self-efficacy also need to be studied to understand a comprehensive picture. Anecdotes of the instructors suggested that some trainees who train for higher ranks often have egos to hold up, which could affect their perception of performance, especially, when comparing themselves to other trainees who are lower in rank. In addition, contextual information about the training could have provided some additional insights. For instance, information about instructors' teaching experience, as this could affect the quality of feedback and the quality of instructions the trainees receive, which in turn affects their learning experience.

Also, task-specific self-efficacy and performance could be studied further in depth. In the present study, task-specific self-efficacy was measured using a newly constructed 10-item questionnaire that reflected general perceived gunner ability as one construct. Single items for general gunner tasks were developed, related to the topics covered in the lessons such as setting up the system, eliminating a static target or moving the target. We chose this approach because we were not aware of the details of the information in the simulator log files until we were granted access to retrieve the log files. Once the log files were retrieved, we found out that in the simulator, the specific gunner tasks are further divided into micro-subtasks. For example, to aim at and follow a target, the gunner first needs to determine the height and the width of the target, or to choose the correct ammunition the gunner needs to identify the type of target and to measure the distance to the target. We did not develop the questionnaire with these detailed micro-tasks in mind. Although this does not mean that our measurement was faulty, we would like to acknowledge that there is scope for further improvement by including additional items to measure self-efficacy for such specific micro-tasks that are trained in the simulator. With such a detailed questionnaire, trainees' perceptions of task-specific self-efficacy can also be related to performance scores for each of these micro-tasks rather than just relying on general number of errors made or an average end score.

Future research needs to consider the effect of qualitative detailed feedback about performance on calibration accuracy compared to the effect of end-of-the lesson quantitative feedback of numerical scores. This could be studied in an experimental set-up in which different groups of trainees receive feedback about their performance, multiple times during the training. This feedback can be either positive, negative, or neutral, or could be provided by different agents, e.g., the instructor or automatically by the simulator. Backlund et al. (2008) showed that positive feedback resulted in a significant increase in self-efficacy than neutral feedback on performance in a game-based driving simulator. Furthermore, Bench et al. (2015) showed that males tend to be overestimating before feedback and become more accurate at estimating their abilities after they receive feedback. Since miscalibrating self-efficacy can have multiple negative impacts on learning and performance (Boekaerts & Rozendaal, 2010; Jernigan, 2004; Moore & Healy, 2008; Vancouver et al., 2002; Zvacek et al., 2015), better insights on whether self-efficacy can be influenced by feedback and what kind of feedback can have a significant impact on self-efficacy beliefs are important.

Another potential avenue for future research could be studying the differences between SBT and real-life training settings. Although from previous research it is known that SBT can be as effective for learning as real-life training (e.g., Bell et al., 2008; Marcano et al., 2019), SBT is never an exact replica of the real-life situation. In that case, performance in a real platform (in this case the CV-90) could be measured. But note that performance on a real platform would still be restricted to a certain extent given that we would not want to eliminate real targets. A large-scale real-life operational training scenario could be suitable for this. It would then be interesting to study whether the relatively lower-stakes SBT environment promotes more overestimation compared to the more high-stake real-life training environment, and how that affects performance.

10 Implications for Practice

The present study has provided indications on the impact of general self-efficacy, task-specific self-efficacy, and (mis)calibration of self-efficacy in a military simulator training. Previous research has shown that miscalibrating self-efficacy can have negative impact on learning and performance (e.g., Boekaerts & Rozendaal, 2010; Jernigan, 2004; Moore & Healy, 2008; Vancouver et al., 2002; Zvacek et al., 2015). It is known that overestimation may lead to risky behavior (Stankov & Lee, 2017), underestimation of the necessary effort (Boekaerts & Rozendaal, 2010; Talsma et al., 2019), lower motivation and less studying (Jernigan, 2004), lower inclination to ask for help (Zvacek et al., 2015), more carelessness in their work (Vancouver et al., 2002), or a lack the crucial realistic basis from which to approach their learning (Moore & Healy, 2008). Our results have shown that the trainees who miscalibrated, tend to overestimate their performance. Furthermore, those who overestimated, performed worse than those trainees who underestimated or were accurate.

These findings are aligned with the existing research reporting how inflated self-efficacy judgements are problematic as it may obscure individuals' perceptions about the efforts required to perform well in a task (e.g., Cleary, 2009). It is important that overestimating trainees are guided towards more realistic expectations. Information about specific (part-task) performance could be provided to the instructors on dashboards to support their feedback. Based on this information instructors could provide trainees with feedback that is better suited to the trainees' needs, both in terms of performance and self-efficacy perceptions. This is important, because carelessness, risky behavior, and overconfidence may increase the risk of making mistakes, which may have detrimental consequences in military operations. Providing trainees with suitable and constructive feedback may result in a more accurate estimation of performance (Bench et al., 2015).

Author contributions *Helena Pennings* was responsible for the conceptualization, data curation, formal analysis, investigation, methodology, project administration, supervision, co-writing the original draft, review and editing of the final draft. *Sindu George* was responsible for co-writing the original draft. *Mathilde Meijer* was responsible for formal analysis, co-writing the original draft.

Funding This research was funded by the Dutch Ministry of Defense [grant number V1806].

Declarations

Conflict of interests The authors have no relevant financial or non-financial interests to disclose.

References

Antonopoulos, P., Fokides, E., & Koutromanos, G. (2024). Understanding learning and learning experience in immersive virtual reality. *Technology, Knowledge and Learning, Advance Online Publication.* https://doi.org/10.1007/s10758-024-09764-z

Backlund, P., Engström, H., Johannesson, M., Lebram, M., & Sjödén, B. (2008). Designing for self-efficacy in a game based simulator: An experimental study and its implications for serious games design. *Paper Presented at the International Conference of Visualisation, London.* https://doi.org/10.1109/VIS.2008.26

Bandura, A. (1978). Self-efficacy: Toward a unifying theory of behavioral change. *Advances in Behaviour Research and Therapy*, *1*(4), 139–161. https://doi.org/10.1016/0146-6402(78)90002-4

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

Bandura, A. (1997). *Self-efficacy: The exercise of control*. W. H. Freeman and Company.

Bandura, A. (2006). Guide for constructing self-efficacy scales. In T. Urdan & F. Pajares (Eds.), *Self-efficacy beliefs of adolescents* (pp. 307–337). IAP.

Barnes, S. L., Bukoski, A., Kerby, J. D., Llerena, L., Armstrong, J. H., Strayhorn, C., Bailey, J., Dorlac, W., Shotto, R., Norfleet, J., Coakley, T., Bowyer, M., Murray, B., Shapiro, M., Manson, R., Moloff, A., Burgess, D., Hester, R., Lewandowski, W., … Honold, E. (2016). Live tissue versus simulation training for emergency procedures: Is simulation ready to replace live tissue? *Surgery, 160*(4), 997–1007. https://doi.org/10.1016/j.surg.2016.04.044

Bell, B. S., Kanar, A. M., & Kozlowski, S. W. (2008). Current issues and future directions in simulation-based training in North America. *The International Journal of Human Resource Management, 19*(8), 1416–1434. https://doi.org/10.1080/09585190802200173

Bench, S. W., Lench, H. C., Liew, J., Miner, K., & Flores, S. A. (2015). Gender gaps in overestimation of math performance. *Sex Roles, 72*(11), 536–546. https://doi.org/10.1007/s11199-015-0486-9

Boekaerts, M., & Rozendaal, J. S. (2010). Using multiple calibration indices in order to capture the complex picture of what affects students' accuracy of feeling of confidence. *Learning and Instruction, 20*(5), 372–382. https://doi.org/10.1016/j.learninstruc.2009.03.002

Bol, L., Hacker, D. J., O'Shea, P., & Allen, D. (2005). The influence of overt practice, achievement level, and explanatory style on calibration accuracy and performance. *The Journal of Experimental Education, 73*(4), 269–290. https://doi.org/10.3200/JEXE.73.4.269-290

Bradley, R. L., Browne, B. L., & Kelley, H. M. (2017). Examining the influence of self-efficacy and self-regulation in online learning. *College Student Journal, 51*(4), 518–530.

Bragard, I., Farhat, N., Seghaye, M. C., Karam, O., Neuschwander, A., Shayan, Y., & Schumacher, K. (2019). Effectiveness of a high-fidelity simulation-based training program in managing cardiac arrhythmias in children: A randomized pilot study. *Pediatric Emergency Care, 35*(6), 412–418.

Chen, G., Gully, S. M., & Eden, D. (2001). Validation of a new general self-efficacy scale. *Organizational Research Methods, 4*(1), 62–83. https://doi.org/10.1177/109442810141004

Chen, P. P. (2003). Exploring the accuracy and predictability of the self-efficacy beliefs of seventh-grade mathematics students. *Learning and Individual Differences, 14*(1), 77–90.

Chiu, M. M., & Klassen, R. M. (2010). Relations of mathematics self-concept and its calibration with mathematics achievement: Cultural differences among fifteen-year-olds in 34 countries. *Learning and Instruction, 20*(1), 2–17. https://doi.org/10.1016/j.learninstruc.2008.11.002

Choi, N. (2005). Self-efficacy and self-concept as predictors of college students' academic performance. Psychology in the Schools. *Psychology in the Schools, 42*(2), 197–205. https://doi.org/10.1002/pits.20048

Cleary, T. J. (2009). Monitoring trends and accuracy of self-efficacy beliefs during interventions: Advantages and potential applications to school-based settings. *Psychology in the Schools., 46*(2), 154–171. https://doi.org/10.1002/pits.20360

Dahlstrom, N., Dekker, S., Van Winsen, R., & Nyce, J. (2009). Fidelity and validity of simulator training. *Theoretical Issues in Ergonomics Science, 10*(4), 305–314. https://doi.org/10.1080/14639220802368864

Davis, W. D., Fedor, D. B., Parsons, C. K., & Herold, D. M. (2000). The development of self-efficacy during aviation training. *Journal of Organizational Behavior, 21*, 857–871.

DeMarree, K. G., Petty, R. E., Bri˜nol, P., & Xia, J. (2020). Documenting individual differences in the propensity to hold attitudes with certainty. *Journal of Personality and Social Psychology, 119*, 1239–1265. https://doi.org/10.1037/pspa0000241

Dixson, D. D., Worrell, F. C., Olszewski-Kubilius, P., & Subotnik, R. F. (2016). Beyond perceived ability: The contribution of psychosocial factors to academic performance. *Annals of the New York Academy of Sciences, 1377*(1), 67–77. https://doi.org/10.1111/nyas.13210

Dunn, K. E., Osborne, C., & Link, H. J. (2014). Research briefs high-fidelity simulation and nursing student self-efficacy: Does training help the little engines know they can? *Nursing Education Perspectives, 35*(6), 403–404. https://doi.org/10.5480/12-1041.1

Dupeyrat, C., Escribe, C., Huet, N., & Régner, I. (2011). Positive biases in self-assessment of mathematics competence, achievement goals, and mathematics performance. *International Journal of Educational Research, 50*(4), 241–250. https://doi.org/10.1016/j.ijer.2011.08.005

Ehrlinger, J., Johnson, K., Banner, M., Dunning, D., & Kruger, J. (2008). Why the unskilled are unaware: Further explorations of (absent) self-insight among the incompetent. *Organizational Behavior and Human Decision Processes, 105*(1), 98–121. https://doi.org/10.1016/j.obhdp.2007.05.002

Fadale, K. L., Tucker, D., Dungan, J., & Sabol, V. (2014). Improving nurses' vasopressor titration skills and self-efficacy via simulation-based learning. *Clinical Simulation in Nursing, 10*(6), 291–299.

Feltz, D. L., Short, S. E., & Sullivan, P. J. (2008). *Self-efficacy in sport*. Human Kinetics. https://doi.org/10.5040/9781718206625

Frank Pajares, M., & Miller, D. (1994). Role of self-efficacy and self-concept beliefs in mathematical problem solving: A path analysis. *Journal of Educational Psychology, 86*(2), 193–203. https://doi.org/10.1037/0022-0663.86.2.193

Gegenfurtner, A., Quesada-Pallarès, C., & Knogler, M. (2014). Digital simulation-based training: A meta-analysis. *British Journal of Educational Technology, 45*(6), 1097–1114.

Gonida, E. N., & Leondari, A. (2011). Patterns of motivation among adolescents with biased and accurate self-efficacy beliefs. *International Journal of Educational Research, 50*(4), 209–220.

Hilpert, J. C., Stempien, J., van der Hoeven Kraft, K. J., & Husman, J. (2013). Evidence for the latent factor structure of the MSLQ: A new conceptualization of an established questionnaire. *SAGE Open, 3*(4), 2158244013510305.

Horcajo, J., Santos, D., & Higuero, G. (2022). The effects of self-efficacy on physical and cognitive performance: An analysis of meta-certainty. *Psychology of Sport and Exercise, 58*, 102063.

Hsu, L. L., Chang, W. H., & Hsieh, S. I. (2015). The effects of scenario-based simulation course training on nurses' communication competence and self-efficacy: A randomized controlled trial. *Journal of Professional Nursing, 31*(1), 37–49.

Huang, X., Mayer, R. E., & Usher, E. L. (2020). Better together: Effects of four self-efficacy-building strategies on online statistical learning. *Contemporary Educational Psychology, 63*, 101924.

Jernigan, C. G. (2004). What do students expect to learn? The role of learner expectancies, beliefs, and attributions for success and failure in student motivation. *Current Issues in Education, 7*(4), 1099–1839.

Johannes, C., & Haase, A. (2024). The impact of feedback mode on learning gain and self-efficacy: A quasi-experimental study. *Active Learning in Higher Education, 25*(3), 391–407. https://doi.org/10.1177/14697874221131970

Kallia, M., & Sentance, S. (2018). Are boys more confident than girls? the role of calibration and students' self-efficacy in programming tasks and computer science. In *Proceedings of the 13th workshop in primary and secondary computing education* (pp. 1–4).

Kizilcec, R. F., Piech, C., & Schneider, E. (2013). Deconstructing disengagement: analyzing learner subpopulations in massive open online courses. *ACM International Conference Proceeding Series, DOI, 10*(1145/2460296), 2460330.

Klassen, R. (2002). A question of calibration: A review of the self-efficacy beliefs of students with learning disabilities. *Learning Disability Quarterly, 25*(2), 88–102. https://doi.org/10.2307/15112

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*, 1121–1134.

Lee, S., & Klein, H. J. (2002). Relationships between conscientiousness, self-efficacy, self-deception, and learning over time. *Journal of Applied Psychology, 87*(6), 1175–1182. https://doi.org/10.1037/0021-9010.87.6.1175

Lee, D., Watson, S. L., & Watson, W. R. (2020). The relationships between self-efficacy, task value, and self-regulated learning strategies in massive open online courses. *International Review of Research in Open and Distributed Learning, 21*(1), 23–39. https://doi.org/10.19173/irrodl.v20i5.4389

Lin, L.-M., & Zabrucky, K. M. (1998). Calibration of comprehension: Research and implications for education and instruction. *Contemporary Educational Psychology, 23*(4), 345–391.

Luctkar-Flude, M., Wilson-Keates, B., & Larocque, M. (2012). Evaluating high-fidelity human simulators and standardized patients in an undergraduate nursing health assessment course. *Nurse Education Today, 32*(4), 448–452.

Maenhout, G., Billiet, V., Sijmons, M., & Beeckman, D. (2021). The effect of repeated high-fidelity in situ simulation-based training on self-efficacy, self-perceived leadership qualities and team performance: A quasi-experimental study in a NICU-setting. *Nurse Education Today, 100*, 104849. https://doi.org/10.1016/j.nedt.2021.104849

Maran, N. J., & Glavin, R. J. (2003). Low-to high-fidelity simulation–a continuum of medical education? *Medical Education, 37*, 22–28. https://doi.org/10.1046/j.1365-2923.37.s1.9.x

Marcano, L., Haugen, F. A., Sannerud, R., & Komulainen, T. (2019). Review of simulator training practices for industrial operators: How can individual simulator training be enabled? *Safety Science, 115*, 414–424. https://doi.org/10.1016/j.ssci.2019.02.019

Marsh, H. W., Pekrun, R., Parker, P. D., Murayama, K., Guo, J., Dicke, T., & Arens, A. K. (2019). The murky distinction between self-concept and self-efficacy: Beware of lurking jingle-jangle fallacies. *Journal of Educational Psychology, 111*(2), 331–353. https://doi.org/10.1037/edu0000281

Miller, T., & Geraci, L. (2011). Unskilled but aware: Reinterpreting overconfidence in low-performing students. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(2), 502e506.

Mitchell, T. R., Hopper, H., Daniels, D., George-Falvy, J., & James, L. R. (1994). Predicting self-efficacy and performance during skill acquisition. *Journal of Applied Psychology, 79*(4), 506. https://doi.org/10.1037/0021-9010.79.4.506

Ministerie van Defensie. (2021). *Jaarverslag en Slotwet Ministerie van Defensie 2020*. Tweede Kamer, 2020–2021, 35 830 X, nr. 1. Den Haag: Unknown.

Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review, 115*(2), 502–517. https://doi.org/10.1037/0033-295X.115.2.502

Morfoot, C., & Stanley, H. (2018). Simulation-based education for neonatal skills training and its impact on self-efficacy in post-registration nurses. *Infant, 14*(2), 77–81.

Multon, K. D., Brown, S. D., & Lent, R. W. (1991). Relation of self-efficacy beliefs to academic outcomes: A meta-analytic investigation. *Journal of Counseling Psychology, 38*(1), 30–38. https://doi.org/10.1037/0022-0167.38.1.30

Nazir, S., Sorensen, L. J., Øvergård, K. I., & Manca, D. (2015). Impact of training methods on Distributed Situation Awareness of industrial operators. *Safety Science, 73*, 136–145. https://doi.org/10.1016/j.ssci.2014.11.015

Nurnindyah, D., Pramadi, A., & Pandjaitan, L. N. (2023). Teachers' Motivational Support, Academic Self-Efficacy and Academic Motivation: The SEM investigation of Naval Cadets' Engagement. *Journal of Educational, Health, Community Psychology, 12*(4), 792–820.

Orvis, K. A., Horn, D. B., & Belanich, J. (2009). An examination of the role individual differences play in videogame-based training. *Military Psychology, 21*, 461–481.

Pajares, F. (1997). Current directions in self-efficacy research. In M. Maehr, & P. R. Pintrich (Eds.), *Advances in motivation and achievement* (Vol. 10, pp. 1e49). Greenwich, CT: JAI Press.

Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research, 66*(4), 543–578. https://doi.org/10.3102/00346543066004543

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

Pajares, F., & Schunk, D. E. (2005). Self-efficacy and self-concept beliefs: Jointly contributing to the quality of human life. In H. Marsh, R. G. Craven, & D. M. McInerney (Eds.), *New frontiers for self research* (pp. 95–122). Information Age Publishing.

Paschke, P., Weidinger, A. F., & Steinmayr, R. (2020). Separating the effects of self-evaluation bias and self-view on grades. *Learning and Individual Differences, 83*, 101940.

Pintrich, P. R., Smith, D. A., Garcia, T., & McKeachie, W. J. (1993). Reliability and predictive validity of the motivated strategies for learning questionnaire (MSLQ). *Educational and Psychological Measurement, 53*(3), 801–813. https://doi.org/10.1177/0013164493053003024

Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin, 138*, 353–387. https://doi.org/10.1037/a0026838

Robbins, S. B., Lauver, K., Le, H., Davis, D., Langley, R., & Carlstrom, A. (2004). Do psychosocial and study skill factors predict college outcomes? *A Meta-Analysis. Psychological Bulletin, 130*(2), 261. https://doi.org/10.1037/0033-2909.130.2.261

Roh, Y. S., Lee, W. S., Chung, H. S., & Park, Y. M. (2013). The effects of simulation-based resuscitation training on nurses' self-efficacy and satisfaction. *Nurse Education Today, 33*(2), 123–128.

Scherbaum, C. A., Cohen-Charash, Y., & Kern, M. J. (2006). Measuring general self-efficacy: A comparison of three measures using item response theory. *Educational and Psychological Measurement, 66*(6), 1047–1063.

Schoenherr, J. R., Waechter, J., & Lee, C. H. (2024). Quantifying expertise in the healthcare professions. In J. R. Scheonherr & M. McConnell (Eds.), *Fundamentals and frontiers of medical education and decision-making: Educational theory and psychological practice* (pp. 213–244). Routledge. https://doi.org/10.4324/9781003316091-10

Schunk, D. H. (1996). *Self-efficacy for learning and performance.* Paper presented at the annual conference of the American educational research association (New York, NY, April 8–12, 1996).

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. https://doi.org/10.2466/pr0.1982.51.2.663

Sitzmann, T. (2011). A meta-analytic examination of the instructional effectiveness of computer-based simulation games. *Personnel Psychology, 64*, 489–528.

Stankov, L., & Lee, J. (2017). Self-beliefs: Strong correlates of mathematics achievement and intelligence. *Intelligence, 61*, 11–16. https://doi.org/10.1016/j.intell.2016.12.001

Stellflug, S. M., & Lowe, N. K. (2018). The effect of high fidelity simulators on knowledge retention and skill self efficacy in pediatric advanced life support courses in a rural state. *Journal of Pediatric Nursing, 39*, 21–26.

Sun, Y., Pan, C., Li, T., & Gan, T. J. (2017). Airway management education: Simulation based training versus non-simulation based training-A systematic review and meta-analyses. *BMC Anesthesiology, 17*(1), 1–7.

Talsma, K., Norris, K., & Schüz, B. (2020). First-year students' academic self-efficacy calibration: Differences by task type, domain specificity, student achievement level, and over time. *Student Success, 11*(2), 109–121. https://doi.org/10.5204/ssj.1677

Talsma, K., Schüz, B., & Norris, K. (2019). Miscalibration of self-efficacy and academic performance: Self-efficacy≠ self-fulfilling prophecy. *Learning and Individual Differences, 69*, 182–195. https://doi.org/10.1016/j.lindif.2018.11.002

Tzamaras, H., Sinz, E., Yang, M., Ng, P., Moore, J., & Miller, S. (2024). Competence over confidence: Uncovering lower self-efficacy for women residents during central venous catheterization training. *BMC Medical Education, 24*(1), 923.

Usher, E. (2016). Personal capability beliefs. In L. Corno & E. M. Anderman (Eds.), *Handbook of educational psychology* (Vol. 3, pp. 146–159). Routledge.

Usher, E. L., Ford, C. J., Li, C. R., & Weidner, B. L. (2019). Sources of math and science self-efficacy in rural Appalachia: A convergent mixed methods study. *Contemporary Educational Psychology, 57*, 32–53.

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

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

Vancouver, J. B., & Kendall, L. N. (2006). When self-efficacy negatively relates to motivation and performance in a learning context. *Journal of Applied Psychology, 91*, 1146–1153. https://doi.org/10.1037/0021-9010.91.5.1146

Vancouver, J. B., Thompson, C. M., Tischner, E. C., & Putka, D. J. (2002). Two studies examining the negative effect of self-efficacy on performance. *Journal of Applied Psychology, 87*(3), 506–516. https://doi.org/10.1037/0021-9010.87.3.506

Whitney, S. J., Fidock, J. J., & Ferguson, N. (2012). Assessing the effectiveness of simulation-based counter-IED training. *Journal of Battlefield Technology, 15*(1), 57–64.

Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. *Contemporary Educational Psychology, 25*(1), 82–91. https://doi.org/10.1006/ceps.1999.1016

Zvacek, S. M., de Fátima Chouzal, M., & Restivo, M. T. (2015). Accuracy of self-assessment among graduate students in mechanical engineering. *Proceedings of the International Conference on Interactive Collaborative Learning (ICL), 2015*, 1130–1133. https://doi.org/10.1109/ICL.2015.7318192