MONITORING ACCURACY ACROSS DOMAINS OF GENERAL KNOWLEDGE AND EMOTIONAL FACE RECOGNITION

WAFA SAOUD

A DISSERTATION SUBMITTED TO
THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN CLINICAL DEVELOPMENTAL PSYCHOLOGY

YORK UNIVERSITY
TORONTO, ONTARIO

MARCH, 2020

© Wafa Saoud, 2020

ABSTRACT

Confidence and its accuracy have been most commonly examined in domains such as general knowledge and learning, and have been rarely studied in social domains. The current study extends the study of metacognitive monitoring accuracy in an academic context, to metacognitive monitoring accuracy in a social context. Monitoring accuracy indices, both calibration and discrimination, were examined for emotional face recognition in 136 university students through an experimental task paradigm developed by Kelly and Metcalfe (2011). To examine whether monitoring accuracy is stable within individuals and across tasks that represent diverse areas, metacognitive monitoring for emotional face recognition was compared to metacognitive monitoring for general knowledge. In addition, correlations between monitoring accuracy and cognitive abilities (intellectual ability and working memory), several aggregated judgments regarding each task as a whole (ratings of predicted and postdictive performance, difficulty, and effort required), as well as self-perceptions relating to social anxiety, empathic ability, differentiation of emotions, and emotion regulation skills were explored. Calibration, but not resolution, was positively correlated across tasks, reflecting a person-centered trait. Cognitive abilities were predictive of monitoring accuracy across tasks, while other task-specific judgments and self-perception variables demonstrated specific associations with monitoring accuracy for emotional face recognition. Monitoring accuracy for emotional face recognition was not predictive of self-reported social-emotional challenges. Overall, study findings support that calibration and resolution are separate indicators of monitoring accuracy, both relevant for understanding metacognitive monitoring in a social domain.

ACKNOWLEGMENTS

I wish to thank all of the people who helped me reach milestones in the completion of this project.

I would like to express my gratitude to my graduate studies supervisor, Dr. Maggie Toplak. You have been a dedicated and dear mentor to me. Your guidance and support was immense to the completion of not only this study, but my whole graduate experience.

I am very grateful for my committee members. Dr. Skilling, thank you for your mentorship throughout my clinical training and this project. You have greatly contributed to my growth as a researcher and future psychologist. Dr. Weiss, I appreciate the opportunity to learn from you both in class and from your expertise on this project.

Thank you also to Dr. Andrew Dane, Dr. Rose Steele, and Dr. Christine Till. I greatly appreciate your contributions and the discussions we shared. I would also like to thank Dr. Rakefet Ackerman, for her insight in this study and helping with methodological challenges.

A special thanks to my lab colleague, Stella Dentakos, for being on my team for this project. It was made richer with our combined efforts and friendship. Amanda Edwards, Cameron Amini, and Jala Rizic, thank you for your contributions.

I would like to recognize the great love that I received from my family and friends throughout my graduate experience. My parents, Antoinette and Faisal, brother, Simon, sister-in law, Viviane, and nephew, Abel, you kept me going.

Finally, thank you, Toronto, for welcoming me to your city with open arms and to all of the friendships I made along the way.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Monitoring accuracy across domains of general knowledge and emotional face recognition

People tend to be poor judges of their own knowledge, believing we know more about a topic than we do (e.g., DeSoete & Roeyers, 2006; Schneider, Visé, Lockl, & Nelson, 2000). High confidence in one's knowledge can be advantageous for decision-making because it may evoke greater ambition, morale, and persistence (see Taylor & Brown, 1988). High confidence in one's knowledge can also be disadvantageous for decision-making because it may lead to lower vigilance and inaccurate assessments of self and others (e.g., Johnson & Fowler, 2011; Dunlosky and Rawson, 2012). The discrepancy between what people know and what people think they know has been examined extensively in both the metacognition field (Kleitman & Stankov, 2007; Koriat, 2008, 2012a; Stankov, Kleitman, & Jackson, 2014) and in the field of judgment and decision-making (Bruine de Bruine, Parker, & Fischoff, 2007; Parker & Fischoff, 2005; Stanovich, West, & Toplak, 2016; Tayes, Lees, & Bush, 1997; West & Stanovich, 1997), showing that we tend to be poor judges of our own knowledge and performance in academic and cognitive domains. However, few studies have compared knowledge and the monitoring of one's knowledge in a social domain. Metacognitive monitoring represents a subjective assessment of how well a task has been, or will be, performed (Ackerman & Thompson, 2017; Bjork, Dunlosky, & Kornell, 2013; Nelson & Narens, 1980). In this study, metacognitive monitoring was examined within a social domain that has been scarcely studied in this context despite being common in real life scenarios: emotional face recognition.

Metacognitive monitoring was examined in the sparsely examined domain of emotional face recognition. I integrated the study of a general knowledge domain to better understand similarities and differences in metacognitive processes across domains. The question of domain-generality versus domain-specificity was explored by examining whether metacognitive monitoring was consistent or variable across these diverse domains. In addition, cognitive abilities, predictive and postdictive confidence judgments, and perceptions of task difficulty and effort were examined as correlates of monitoring accuracy in both domains. Various self-perceptions and social-emotional dispositions were further explored as correlates of monitoring accuracy in an emotional face recognition domain specifically. Monitoring accuracy in an emotional face recognition domain was explored in relation to a behavioural correlate: social-emotional challenges in university students.

Monitoring Accuracy: Definition and Measurement

Metacognition has been defined as "thinking about thinking" (Flavell, 1979) and includes two primary components: metacognitive monitoring and metacognitive control (Schraw & Dennison, 1994). Metacognitive monitoring represents the ability to evaluate current states of knowledge and skills, such as assessing the correctness of one's own performance on a task. Metacognitive control represents the ability to engage in regulation strategies, such as double-checking one's understanding or seeking help (Grainger, Williams, & Lind, 2016; Zabrucky, 2010; Ackerman, 2019). Metacognitive monitoring is believed to precede metacognitive control and to guide effort regulation (Zabrucky, 2010). The accuracy of metacognitive monitoring can be measured in two ways: calibration and resolution.

Calibration. Calibration represents the degree of fit between subjective feelings of certainty or correctness about one's response, known as confidence judgments, and the objective accuracy of one's performance on a task (Keren, 1991). The more closely subjective confidence judgments match objective performance, the better calibrated the individual is considered. Calibration is thought to be a measure of *bias*, measuring how precisely confidence matches performance (Baranski & Petrusic, 2009) and is, thus, a measure of absolute accuracy.

Poor calibration can happen in two directions: overconfidence and underconfidence (Lichstenstein & Fischoff, 1977). Overconfidence occurs when subjective confidence judgments are greater than objective performance, reflecting a failure to detect errors (e.g., confidence in incorrect responses; Pallier, Wilkinson, Danthir, Kleitman, Knezevic, et al., 2002; Rinne & Mazzocco, 2014). In contrast, underconfidence occurs when subjective confidence judgments are lower than actual performance, reflecting the false detection of errors (e.g., a lack of confidence in correct responses; Pallier et al., 2002; Rinne & Mazzocco, 2014). The following example illustrates the difference between underconfidence and overconfidence: two patients have a heart condition. A doctor may correctly diagnose one of these two patients with a heart condition, and this doctor may be 100% confidence that they detected all heart conditions. However, this doctor missed the second patient with the heart condition. Thus, this doctor would be considered overconfident. In contrast, this doctor may diagnose both patients with a heart condition, and be 50% confident that they detected all heart conditions. This doctor did detect all heart conditions and, thus, would be considered underconfident.

Subjective confidence has an impact on reasoning and decision-making, by regulating and directing behaviors (e.g., See Livingston, 2003; Dunlosky & Rawson, 2012; Hacker, Bol & Keener, 2008b). Overconfidence can lead to a false sense of mastery, resulting in less cognitive resource investment than needed to understand a situation or solve a problem. For example, if knowledge and skills in a given situation is judged as satisfactory, an individual may not take further steps to improve their performance. In contrast, underconfidence can lead to continued, and perhaps unnecessary, investment of cognitive resources to a task or problem. For example, if knowledge and skills in a given situation is judged as inadequate, an individual may seek additional information to improve their understanding and decision-making. Therefore, well-developed calibration skills are critical for regulating resource allocation in a situation.

However, individuals have been shown to be poor judges of their own knowledge state (Koriat, Lichtenstein & Fischoff, 1980; Yates, Lee & Bush, 1997; Sporer, Penrod, Read & Cutler, 1995). A tendency towards overconfidence, rather than underconfidence, has been displayed by both children and adults (Bjork et al., 2013; Soderstrom, Yue, & Bjork, 2015; Stanovich, West, & Toplak, 2016). For example, educational psychologists have been interested in studying monitoring accuracy as it relates to academic functioning, and have found that it plays a critical role in successful learning (See Livingston, 2003; Dunlosky & Rawson, 2012). Students tend to be more confident than accurate in their performance on tests of mathematical and general knowledge, subsequently impacting on effort regulation (Ackerman & Zalmanov, 2012; Prowse, Turner & Thompson, 2009; Shynkaruk & Thompson, 2006). For another example, forensic psychologists have studied confidence judgments in eye-witnesses. Eye-witnesses are not always accurate at identifying perpetrators, and they tend to express high confidence in their identification, even when they are wrong (Sporer et al.,1995; Wells, Olsen & Charman, 2002). A tendency towards overconfidence has also been demonstrated cross-culturally. For example, although overconfidence in general knowledge has been demonstrated to be generally stronger in Asian participants than Western participants (Yates et al.,1997), both of these populations display a general tendency towards overestimating general knowledge.

When a problem is complex or ambiguous, accurate monitoring of our knowledge state is especially important because it may increase the likelihood of more closely assessing our comprehension and disconfirming erroneous beliefs (see Thompson, Turner & Pennycook, 2011; Thompson, Turner, Pennycook & Bol, 2013; Zimmerman & Moylan, 2009). Many social situations are ambiguous or complex and require interpretation. Biased interpretations of social information can contribute to problems with emotional and social functioning (e.g., Dodge & Crick, 1990; Dodge, Lansford, Burks, Bates, Pettit et al.,2003; Lemerise & Arsenio, 2000; de Castro, Merk, Koops, Veerman, & Bosch, 2005; Penn, Mueser, Spaulding, Hope, & Reed, 1995; Clark & Wells, 1995; Amir, Foam & Coles, 1998). Confidence has been found to predict information-seeking in decision-making contexts (Desender, Boldt, & Yeung, 2018). Relevant to social contexts and according to metacognitive theory (see Ackerman, 2019), if our understanding of a social situation is considered to be uncertain, we might be more likely to seek additional information to clarify our understanding, consequently disconfirming biased beliefs. In contrast, if our understanding of a social situation is considered to be satisfactory, we might be less likely to seek additional information. Therefore, confidence may play an important role in guiding individuals to invest less or more effort not only in academic situations as has been previously shown in the scientific literature, but also in social situations.

In the current study, the measurement of confidence judgments was based on an experimental paradigm that is commonly used in the metacognition literature (Ronis & Yates, 1987; Yates, Zhu, Ronis & Wang, 1989; Was, 2014; Parker & Fischhoff, 2005). In this paradigm, participants complete a task and make judgments about their confidence in the accuracy of their answers for each item completed. Confidence judgments are made on a continuous scale. This paradigm allows for direct comparison between mean confidence ratings and mean success for a given task. To measure calibration in the current study, the following indices were used: (1) Bias Index (2) Overconfidence index and (3) Resistance to Overconfidence Index (see Method section for detailed description of calibration indices and corresponding formulas).

An Index of Bias. The Bias Index captures the degree of underconfidence or overconfidence on a test (Ronis & Yates, 1987; Yates et al.,1989; Was, 2014). It is the difference between the mean subjective probability of a correct answer (i.e., mean confidence rating) and the proportion of correct answers (i.e., percent of items answered correctly). Underconfidence occurs when confidence judgments are lower than actual performance; overconfidence occurs when confidence judgments are greater than actual performance. Values range from -1 to +1. Values below zero indicate underconfidence and values above zero indicate overconfidence. For example, if on a given task an individual has a mean confidence rating of 100% (i.e., 1.00) but only answered 50% of the questions correctly (i.e., 0.50), then this individual is considered to be overconfident. In contrast, if on a given task an individual has a mean confidence rating of 50% (i.e., 0.50) but answered 100% of the questions correctly (i.e., 1.00), then this individual is considered to be underconfident. If an individual has a mean confidence rating of 50% (i.e., 0.50) and answered 50% of the questions correctly (i.e., 0.50), then this individual is considered to have no calibration bias, because confidence was perfectly matched to performance. The current study used the Bias Index as one index of calibration.

An Index of Overconfidence. As both ends of the Bias Index measure a type of miscalibration, the Bias Index cannot be meaningfully used in correlational analyses with other individual differences variables. The discrepancy between one's confidence and one's performance on a test can also be captured by taking the absolute difference between mean confidence and mean performance, which is an index of overconfidence (Bruine de Bruin et al., 2007). Values on the overconfidence index range from 0 to 1. Higher scores represent greater discrepancy between performance and confidence and lower scores represent lower discrepancy between performance and confidence. For example, if on a given task one individual has a mean confidence rating of 50% (i.e., .50) and answers 100% of the questions correctly (i.e., 1.00) and another individual has a mean confidence rating of 100% (i.e., 1.00) and answers 50% (i.e., .50) of the questions correctly, both individuals would be considered to have an overconfidence score of .50, indicating poor correspondence between confidence and accuracy. In the research literature, the tendency is for individuals to display overconfidence rather than underconfidence (Stanovich, West & Toplak, 2016). The absolute discrepancy between confidence and accuracy was referred to as the overconfidence index. However, this index measures the magnitude of miscalibration, rather than the direction of miscalibration.

An Index of Resistance to Overconfidence. The Resistance to Overconfidence Index was also used in the correlation and regression analyses, so that a higher score indicates better calibration. Resistance to overconfidence was derived by subtracting the overconfidence index from 1. The Resistance to Overconfidence Index was also consistent the index of resolution, where a higher score indicates better resolution.

Resolution. Resolution, also known as discrimination, is another index of metacognitive monitoring. Resolution indicators are used to assess whether an individual is able to discriminate between correct and incorrect responses (Koriat, 2012a). They are viewed as a measure of *sensitivity* (Koriat, 2012a; 2009). Resolution is considered to be a measure of relative, rather than absolute, accuracy because it assesses confidence for correct items relative to confidence for incorrect items.

Resolution helps guide individuals to effectively choose which aspects of their performance to invest additional effort in to make the best use of their time (e.g., Destan & Roebers, 2015; Thiede, Anderson, & Therriault, 2003). Calibration measures whether we are overestimating our performance in a domain, but resolution measures whether we can discriminate between which parts of our performance in a domain are satisfactory and which parts are inadequate and in need of additional effort investment. For example, you may generally overestimate your understanding of social situations (poor calibration), but recognize that your understanding for some types of social situations is poor and that your understanding for other social situations is satisfactory (good resolution). The following example is provided to illustrate the difference between calibration and resolution: an individual may estimate that a coin will be flipped with .50 probability that it will land on tails. With repeated events, this individual would be perfectly calibrated because coins will hypothetically land on tails half the time. However, this individual would not have good resolution because this individual did not discriminate between situations when the coin landed on heads and when the coin landed on tails. As evidenced by these examples, calibration and resolution measures are dissociable. That is, in the same experimental paradigm, calibration might be high, while resolution might be low, or vice versa (Maki, Shields, Wheeler, & Zacchilli, 2005; see Thiede, Mueller, & Dunlosky, 2015, for a review).

There are many ways to measure resolution (see Flemming & Lau 2014, Nelson, 1984; Schraw, 2009; Jackson & Kleitman, 2014). Similar to other studies within the educational field as well as the decision-making field, the current study measured resolution with the Goodman–Kruskal Gamma correlation. Gamma correlations measure, within-participants, whether correct responses are also the responses that receive higher confidence judgments and whether incorrect responses are also the responses that receive relatively lower confidence judgments (Maki et al., 2005). A gamma value of “zero” indicates a lack of relationship (i.e., lack of discrimination) and greater values indicate stronger relationships (i.e., higher confidence for correct answers and lower confidence for incorrect answers; see Method section for detailed description of the resolution index).

Importantly, calibration and resolution represent different aspects of monitoring accuracy and they have different functions. These indices are dissociable and in the same experimental paradigm, calibration may be high while resolution may be low, or vice versa. (Koriat, Sheffer, & Ma'ayan, 2002). However, performance monitoring may not only vary across different measurement methods, but also across different knowledge domains (Erickson & Heit, 2015; West & Stanovich, 1997). Therefore, the current study included both aspects of monitoring accuracy- calibration and resolution- with a focus on extending monitoring accuracy research to a social domain.

Does Domain Matter?

Overconfidence has been displayed across various domains, including predictions of sports outcomes (Ronis & Yates, 1987), reading comprehension (Glenberg & Epstein, 1987; Lin & Zabrucky, 1998), problem solving (Ackerman & Zalmanov, 2012; García et al., 2016), financial decision-making (Malmendier & Tate, 2008; Zacharakis & Shephard, 2001; Schrand & Zechman, 2012) general knowledge (Koriat, Lichtenstein, & Fischhoff, 1980; Yates, Lee, & Bush, 1997), eyewitness memory (Sporer et al.,1995; Wells, Olsen & Charman, 2002) and emotional face recognition in clinical populations (Grainger, Williams, & Lid., 2016). Some domains are characterized by poor resolution (e.g., reading comprehension; Thiede, Griffin, Wiley, & Anderson, 2010). Others are characterized with high resolution (e.g., problem solving; Ackerman & Zalmanov, 2012; emotional face recognition; Kelly and Metcalfe, 2011; financial calculations; Dentakos, Saoud, Toplak, & Ackerman; 2019). Despite the breadth of studied domains, these domains have been explored mainly in isolation and rarely in parallel (Erickson

& Heit, 2015). Moreover, social functioning has been a largely unexplored domain. Therefore, it is important to consider whether calibration and resolution represent abilities that are general and independent of the domain under study, or whether they are specific and dependent on content knowledge and therefore variable across areas of functioning.

The domain-general hypothesis proposes that one's ability to endorse confidence judgments that precisely match one's performance reflects a skill, or trait, that one consistently applies across different domains (e.g., Erickson & Heit, 2015; Kleitman; 2008; Kleitman & Stankov, 2001; Pallier et al., 2002; Veenman & Verheij, 2003). There has been no evidence in the scientific literature to suggest that resolution is domain-general. However, overconfidence has been emerging in the scientific literature, across many empirical studies, as a general characteristic of individuals. For example, Jackson and Kleitman (2014) found that confidence judgments were similarly biased on a medical decision task and cognitive ability measures (Raven Matrices and a vocabulary test). For another example, in a study by Schraw, Dunkle, Bendixen, & Roedel (1995), students' monitoring accuracy scores were found to be correlated across five different subject domains. Therefore, participants who tended to be overconfident on one type of task in these studies, tended to be overconfident on other types of tasks. According to the domain-general hypothesis, being more confident than accurate is a pervasive bias that an individual will exhibit consistently across all domains of functioning.

In contrast, the domain-specific hypothesis proposes that calibration involves the ability to assess content knowledge that is specific to a given domain. According to the domain-specific hypothesis, individuals' monitoring abilities can vary depending on the domain being considered. Studies in support of the domain-specific hypothesis have demonstrated that calibration indictors are inconsistent across different areas of functioning, suggesting that they are sensitive to domain content (e.g., Glaser, 1991).

Little is known about monitoring accuracy in a social domain and about how monitoring accuracy may vary within individuals from an academic domain to a social domain. To gain an understanding of monitoring accuracy in a social domain and contribute to the understanding of performance monitoring across domains, calibration and resolution were examined in the well-studied domain of general knowledge (Lichtenstein, Fischhoff, & Phillips, 1977; Yates et al.,1997) and compared to an understudied domain in this context: emotional face recognition (Kelly & Metcalfe, 2011). For each of those domains, the research has focused on either resolution or calibration, but no study to my knowledge has examined both measures and none have compared these measures within individuals across these particular domains.

Social Domain. Researchers have begun to examine monitoring accuracy in a social domain, but this research is sparse. Despite the complexity of our social worlds, monitoring accuracy in a social domain has been understudied and is thus a central focus of consideration in the current project.

Emotional face recognition. Important to social functioning is the ability to understand what another person is feeling (Lemerise & Arsenio, 2000; Nelson & Perry, 2015). Knowing another person's emotional state is a critical step in avoiding conflict and providing social support. If people are accurate in knowing what they know about what another person is feeling, then they do not need to devote more time and effort to seeking additional information. If people are accurate in knowing what they *do not* know about what another person is feeling, they can devote more time and effort to appropriately seek and clarify information. Recognizing emotional expressions can be measured for accuracy, and accurate recognition serves a critical function within interpersonal interactions, providing information about another's underlying mental states, their needs, and their likely behaviors (Lemerise & Arsenio, 2000; Nelson & Perry, 2015). Emotion recognition is the process of identifying what another is feeling, usually based on their facial expressions. It is typically assessed using a laboratory paradigm that presents participants with images of facial expressions (e.g., happy; disgusted) and asks participants to identify each facial expression. Emotion recognition is found to be an automatic, universal process (e.g., Ekman, 1972), but the ability to identify emotions accurately fluctuates across individuals (e.g., Kohler et al., 2003; Elfenbein & Ambady, 2002; Adolph, Sears, & Piven, 2001). The ability to interpret emotional expressions is perhaps an important skill at any age, but it may become even more important in emerging adulthood, as goals related to building rewarding and close interpersonal relationships are considered (Carstensen, Isaaco-witz, & Charles, 1999; Salmela-Aro, Aunola, & Nurmi, 2007). However, because research on the metacognitive monitoring of emotion recognition has been sparse, I chose to examine emotional face recognition in the current study in a sample of undergraduate students.

Only one study is known to have examined resolution in emotional face recognition, concluding that individuals can discriminate between facial expressions they know and facial expressions they do not know (Kelly & Metcalfe, 2011). Kelly and Metcalfe (2011) measured emotion recognition through tasks that required participants to identify emotional expressions and obtained self-ratings of confidence for each choice. According to the Goodman-Kruskal Gamma correlation for an index of resolution, participants had higher confidence for correct answers and lower confidence for incorrect answers, indicating that they were good at judging when they know and do not know an emotional expression. It was concluded that people possess good awareness in their ability to understand another's emotional expressions. However, to my knowledge, no other study has attempted to replicate this experiment in the scientific literature.

Additionally, to my knowledge, no studies have examined calibration for emotional face recognition. As described above, calibration and resolution are thought to be separate indices of monitoring accuracy and resolution may be high, while calibration might be low, or vice versa (Koriat, Sheffer, & Ma'ayan, 2002). Therefore, individuals may have high resolution (i.e., higher confidence for correct choices than incorrect choices) but low calibration (i.e., still be overconfident in their answers overall). For example, people may be able to confidently identify when a certain facial expression has occurred relative to a different facial expression (i.e., high resolution), but still believing overall that they are better at identifying facial expressions than they are (i.e., low calibration). Therefore, an important question to explore is whether or not people have an overconfidence bias for emotional face recognition.

A few studies have empirically examined differences in confidence and accuracy for emotional face recognition between clinical and non-clinical populations, generally demonstrating greater variability in confidence than accuracy between populations. McMahon, Henderson, Newall, Jaime, and Mundy (2016) examined differences in confidence and accuracy between children with and without autism spectrum disorder (ASD). Participants viewed pictures of facial expressions (taken from the Reading the Mind in the Eyes task; Baron-Cohen, 2001), identified the affective expression of each face, and rated their confidence for each choice. Although accuracy did not differ significantly between groups, children with ASD were significantly more confident in their choices than children without ASD. Schilling, Wingenfeld, Lowe, Moritz, Terfehr, Kother, and Spitzer (2012) similarly examined differences in confidence and accuracy for emotional face recognition on the Reading the Mind in the Eyes task (Baron-

Cohen, 2001) between individuals with and without Borderline Personality Disorder (BPD). Although accuracy for emotional face recognition did not differ significantly between groups, participants with BPD were significantly more confident in their choices than participants without BPD. However, these studies did not compare confidence and accuracy within individuals across domains.

General Knowledge Domain. General knowledge refers to accumulated knowledge about several topics, as opposed to in-depth knowledge about one specific topic. In the metacognition literature, general knowledge questions have been extensively used in experimental paradigms to examine calibration (Lichtenstein et al., 1977; Yates et al., 1997). In general, participants tend to display overconfidence on general knowledge questions (Bruine de Bruin et al., 2007; Stanovich et al.,2016; West & Stanovich, 1997; Yates et al., 1997). Therefore, overconfidence in the domain of general knowledge reflects a well-established calibration bias in the metacognition literature.

One factor that complicates a comparison of confidence judgments across different domains is item difficulty (Koriat, 2012a). In what is termed the hard-easy effect, difficult test items tend to result in low accuracy and overconfidence, but easy test items tend to result in high accuracy and underconfidence (Justlin, Winman, & Olsson, 2000; Lichtenstein & Fischhoff, 1977). Similarly, individuals with low knowledge in a given area have been shown to be more overconfident than individuals with high knowledge in that area (see Miller & Geraci, 2011). These findings suggest that it is important to use tasks comprised of items with similar difficulty levels when comparing monitoring abilities across domains (e.g., Erickson & Heit, 2015). For example, calibration in general knowledge cannot be fairly compared to calibration in emotional face recognition if the test items in each domain are not matched for difficulty. Therefore, the experimental tasks developed for the current study were piloted in order to try to match item difficulty across general knowledge and emotion recognition domains.

The current study examined monitoring accuracy, using both calibration and resolution indices, within and across social and general knowledge domains. As has been reported in the literature, overconfidence was expected in emotional face recognition and general knowledge domains, consistent with studies showing overconfidence bias in other domains (Bjork et al., 2013; Soderstrom et al., 2015; Stanovich et al., 2016). Therefore, it was hypothesized that overconfidence, measured by the Bias Index, would occur in both domains. Specifically, as Bias scores greater than "0" indicate greater overconfidence, it was hypothesized that Bias scores would be significantly greater than "0" in both domains. As has been reported by Kelly and Metcalfe (2011), discrimination between correct and incorrect answers was expected in an emotional face recognition domain. Similarly, discrimination between correct and incorrect answers was expected for general knowledge. Therefore, it was hypothesized that resolution, measured by the Goodman-Kruskal Gamma correlation, would occur in both domains. Specifically, as resolution scores greater than "0" indicate better discrimination, it was hypothesized that resolution scores would be greater "0" in both domains. It was hypothesized that calibration and resolution would not be correlated within domains, thus informing on different aspects of monitoring accuracy. Consistent with previous research showing calibration is domain-general (Jackson & Kleitman, 2014; Schraw et al., 1995) it was hypothesized that calibration for general knowledge would be correlated with calibration for emotional face recognition. In contrast, consistent with previous research suggesting resolution is domain-specific (Jackson et al., 2016), it was hypothesized that resolution for general knowledge would not be correlated with resolution for emotional face recognition. Table 1 includes a description of the main variables related to monitoring accuracy examined in the current study.

Table 1.
Definitions of Main Terms and Variables

Variable	**Definition**
Metacognition	
Accuracy/ Success Rate	Accuracy represents task performance. It is measured by whether items are successfully completed.
Confidence	A confidence judgment represents a subjective assessment about the correctness of one's answer. Confidence was measured through a confidence rating scale (see methods).
Monitoring accuracy	Monitoring accuracy represents a subjective assessment of task performance. Monitoring accuracy was measured by calibration and resolution indices.
Calibration	
Calibration	Calibration represents the degree of fit between confidence judgments and objective accuracy. Calibration is measured by the Bias index, the Confidence index, and the Resistance to Overconfidence index.
Bias Index	The Bias index represents the direction of miscalibration. Underconfidence occurs when confidence judgments are lower than actual performance. Overconfidence occurs when confidence judgments are greater than actual performance.
Overconfidence Index	The Overconfidence index Represents the magnitude of miscalibration. It is the absolute discrepancy between subjective confidence and objective accuracy. Higher scores indicate poorer calibration.
Resistance to Overconfidence Index	The Resistance to Overconfidence index is the inverse of the Overconfidence Index. It also represents the discrepancy between subjective confidence and objective accuracy and assesses the magnitude of poor calibration. Higher scores indicate better calibration.
Resolution	
Resolution Index	The resolution index represents the degree of discrimination between correct and incorrect items. It is measured by the Goodman-Kruskal Gamma correlation (*G*). This score represents

	the mean within-participant correlation between subjective confidence and objective accuracy. Higher scores indicate better resolution.

Correlates of Monitoring Accuracy

Relatively little is known about the role of individual and dispositional factors on confidence judgments in reasoning and problem-solving contexts. A common individual difference variable studied in the overconfidence literature has been cognitive abilities (Thompson, 2009; West & Stanovich, 1997; see Ackerman & Thompson, 2017, for a review). Various factors are also thought to influence metacognitive judgments, such as task-specific characteristics and person-level characteristics (Efklides, 2011; see Ackerman, 2019). Ackerman (2019) identified three levels of cues that may impact confidence judgments: (1) self-perceptions (e.g., perceived domain knowledge; anxiety "I am good/bad in this domain"), (2) task characteristics (e.g., format; difficulty), and (3) momentary experiences (e.g., familiarity with content). Although affect has also been suggested to impact on metacognitive monitoring (Efklides, 2011), no known studies have examined whether the regulation of one's affective experiences plays a role in monitoring accuracy. To extend on the existing literature on individual differences in monitoring accuracy, in the current study variability of confidence judgments were explored not only across emotional face recognition and general knowledge domains, but also in relation to cognitive abilities, task-specific judgments (i.e., predictive and postdictive confidence; difficulty; effort), certain self-perceptions (i.e., empathic ability; social anxiety; emotion differentiation) and cognitive emotion regulation. Efklides (2011) proposes that when individuals are approaching a task, they build a framework based on a task analysis. For example, on an emotional face recognition task, an individual may think "this is a face with an emotional expression"; "I know how to read emotions" (i.e., self-perception of domain knowledge); “These types of problems are easy” (i.e., task-specific perception of difficulty); “I like these types of problems” (i.e., affective reaction); “I can solve this problem” (i.e., task-specific predictive confidence). Such task-specific judgments and person-level characteristics may impact on performance and performance monitoring.

Cognitive abilities. Two indicators of cognitive abilities have been the main focus in research exploring person-related differences: intelligence and executive functions (Stanovich & West, 1998). Intelligence is usually measured through the assessment of both crystallized and fluid abilities, and executive functions have been indexed by measures of working memory abilities (Komori, 2016). Research has found that general knowledge is positively associated with intelligence (Stanovich & West, 1998; Peters & Bjalkebring, 2015) and working memory (Komori, 2016) in adults. Similarly, performance on emotion recognition tasks has been found to be positively associated with performance on working memory tasks (e.g., Hoaken, Allaby, & Earle, 2007; Phillips, Channon, Tunstall, Hedenstrom, & Lyons, 2008), and intelligence tests (e.g., in children, McAlpine, Kendall, Singh, 1991; in adults, Simon, Rosen, Grossman, & Pratowski, 1995).

Performance on knowledge calibration paradigms, such as those showing overconfidence, has also been associated with cognitive abilities, such as verbal and nonverbal intellectual abilities (Bruine de Bruin et al.,2007; Stanovich & West, 1998; Stanovich et al., 2016) as well as self-reported SAT scores (Stanovich, West, & Toplak, 2016). Therefore, better calibration has been shown to be associated with higher cognitive abilities. This association has been attributed to the simulating and hypothetical reasoning that is required for successful performance on various decision-making tasks (Stanovich, 2011). Cognitive decoupling mechanisms allow individuals to simulate alternative worlds and to consider hypothetical scenarios that are not in their immediate environment, requiring the engagement of analytic processes to construct these scenarios. If poor calibration results from individuals' exposure to environmentally unrepresentative knowledge, then it is possible that at least some cognitive decoupling would be needed to achieve better calibration (West & Stanovich, 1997). Cognitive decoupling has been indexed by intellectual abilities and executive functions (Stanovich, 2009). Therefore, due to cognitive decoupling mechanisms, it was expected that cognitive abilities would be positively correlated with calibration in the current study, replicating previous research (Bruine de Bruin et al., 2007; Stanovich & West, 1998).

Resolution has also been associated with cognitive abilities. Specifically, resolution has been related to depth of processing on reading comprehension tasks (Thiede et al., 2003). Studies that have examined methods to encourage greater depth of processing in readers have shown that greater processing improves resolution (Anderson & Thiede, 2008; Fukaya, 2013; Thiede, Dunlosky, Griffin, & Wiley, 2005; Thiede, Wiley, & Griffin, 2011). Calibration of reading comprehension and problem solving was also found to be improved by tasks that require increased depth of processing (Lauterman & Ackerman, 2014; Sidi, Spigelman, Zalmanov, & Ackerman, 2017). Thus, resolution for both general knowledge and emotional face recognition were also expected to be positively associated with cognitive abilities in the current study.

Cognitive ability, indexed by measures of intellectual ability and working memory, was examined in the current study. It was hypothesized that performance on tests of cognitive ability would be positively correlated with resolution and calibration in both general knowledge and emotional face recognition domains.

Task-specific perceptions. In the metacognition literature, task-specific judgments represent the beliefs individuals have about factors that impact performance on a specific task

(Ackerman, 2019). These beliefs can pertain to task-related abilities, task difficulty, and the amount of effort required to complete the task (Ackerman, 2019). To gain a better understanding of relationships between task-specific perceptions and performance monitoring, participants in the current study were asked to provide several aggregate judgments, including predictive and postdictive confidence judgments, and ratings of task difficulty, required effort, and discomfort.

***Predictive and postdictive aggregate confidence judgments*.** Predictive aggregate confidence judgments represent one's subjective confidence in one's knowledge or abilities, before completing a task. Postdictive aggregate judgments represent one's subjective confidence in overall performance, after completing a task. Predictive confidence judgments require individuals to estimate how they will perform on a task based on stored memory structures relevant to a given domain and it has been suggested that these judgments guide regulation of mental resources on the task (see Ackerman & Thompson, 2017). For example, if an individual estimates high confidence in their ability to complete a task, then they may show increased persistence and effort when faced with challenging questions, believing that this will lead to success. Predictive judgments are considered to be theory-based, guided by one's own beliefs about one's abilities in a certain domain (Ackerman, 2019; Koriat, 1997). In contrast, postdictive confidence judgments require individuals to use their perceived performance on a task they just completed as a reference point for their rating. Therefore, postdictive judgments are considered to be experience-based, guided from one's real-time experience of having recently completed the task (Ackerman, 2019; Koriat, 1997). Predictive and postdictive confidence judgments of aggregate performance have been shown to be positively correlated with success rates in different domains (Erickson & Heit, 2015). In the current study, it was expected that predictive and postdictive confidence would be positively correlated with accuracy and confidence, consistent with previous research (Erickson & Heit, 2015), and positively correlated with calibration and resolution in domains of general knowledge and emotional face recognition.

Task difficulty, effort, and experienced effort. Following each experimental task, participants were also asked to rate task difficulty, effort required, and their affective reaction to engaging in effort (Finn, 2010; Hsu, Eastwood, & Toplak, 2017; Hsu, Propp, Panetta, Martin, Dentakos et al.2018).

Research has shown that task difficulty impacts confidence ratings (Juslin et al., 2000; Klayman & Soll 1999). As described above for the "hard-easy effect", harder items tend to evoke overconfidence, while easier items tend to evoke underconfidence (Juslin et al., 2000). Therefore, it was expected that task difficulty would be negatively correlated with monitoring accuracy in the current study. Specifically, it was hypothesized that higher task difficulty ratings would be correlated with poorer calibration and resolution scores.

Effort, and the affective experience accompanying effort, has received little empirical study in relation to monitoring accuracy. Studies that have explored mental effort in relation to confidence judgments have yielded mixed findings. Some research has shown that tasks perceived as effortful receive lower confidence judgments (e.g., "I worked very hard, so I must be doing it wrong"; for review, see Ackerman 2019). Other research has shown that the experience of increased effort exertion leads to higher subjective confidence judgments (e.g., "I worked very hard so I must be right"; Chen, 2002). In a separate line of research, individuals high in mental effort tolerance have displayed lower difficulty ratings (Dornic, Ekehammar, & Laaksonen, 1991). This finding suggests that individuals who judge a task as more difficult, are also more likely to have a lower ability to tolerate the aversiveness of increased effort expenditure. In a study by Hsu et al. (2018), effort invested was experimentally distinguished from the affective experience of effort invested, and it was shown that difficulty and discomfort were positively correlated, but different from one another. Due to the relationship between increased difficulty and overconfidence described for the hard-easy effect, it was hypothesized that poorer monitoring accuracy would be correlated with increased perceptions of task difficulty, as well as increased effort expenditure and discomfort.

In the current study, participants were asked to provide a rating of predictive confidence, postdictive confidence, task difficulty, required effort, and experienced effort for each task. It was important that these ratings were separated for each task so that participants had a specific reference point for their ratings in each domain. Relationships between monitoring accuracy indices (i.e., calibration and resolution) and these additional aggregate judgments were examined for each task. Overall, it was hypothesized that better calibration and resolution scores would be correlated with higher predictive and postdictive confidence and lower task difficulty, effort, and experienced effort ratings for both general knowledge and emotional face recognition.

Self-perceptions. Self-perceptions include beliefs about one's abilities (e.g., "I am empathetic"), knowledge (e.g., level of expertise), and personal characteristics (e.g., "I am anxious with public speaking"), and relate to one's confidence in one's ability to perform well in a given area (Ackerman, 2019; Jiang & Kletman, 2014; Shcraw & Dennison, 1994). The impact of self-perceptions on monitoring abilities and indices, such as calibration and resolution, have been sparsely explored in metacognition research and have not been explored in relation to metacognitive monitoring in a social domain. As central to the current project is extending the literature on metacognitive monitoring in a social domain, self-perceptions of empathic ability, emotion differentiation, and social anxiety were explored as potential monitoring accuracy correlates for emotional face recognition.

Empathy. Kröner and Biermann (2007) suggested that when individuals are asked to form a confidence judgment on a task, they are likely to base their confidence judgments, in part, on how they perceive themselves to perform on similar tasks. Ackerman (2019) suggests that this process serves as a cue to help form confidence judgments. An individual's self-perceptions can vary according to context (e.g., academic self-perceptions versus social self-perceptions). Therefore, measures of self-perceptions should be related to the domain being considered. Empathy is defined as an affective reaction that is elicited by the apprehension or comprehension of another's emotional state (Eisenberg & Fabes, 1998). Realo et al. (2003) found that self-reported empathy failed to predict performance on a face or voice emotion recognition task. Similarly, Ickes, Stinson, Bissonnette, and Garcia (1990) found that while participants varied in their ability to recognize the emotions exchanged during an unstructured dyadic interaction, individuals who self-reported as having skill at understanding another's emotions were not more accurate at identifying emotions than those who reported less self-competence. While perceived empathic ability has been found to be unrelated to emotion recognition performance (Relo et al., 2013; Ickes, Stinson, Bissonnette, & Garcia; 1990), perceived empathic ability may influence one's confidence on a test of emotional face recognition, and serve as a cue for monitoring one's performance. Therefore, perceived empathic ability was considered in the current study. It was hypothesized that perceived empathic ability would be correlated with increased confidence and better calibration and resolution for emotional face recognition, in line with studies showing that beliefs about one's academic skills relates to monitoring accuracy in an academic domain (Ackerman, 2017; Jiang & Kletman, 2014; Shcraw & Dennison, 1994).

Social Anxiety. Anxiety can also vary depending on the context, and measures of anxiety should be related to the domain being considered (see Jiang & Kleitman, 2014). Given the current study's focus on examining metacognitive accuracy within a social domain, social anxiety was explored. Social anxiety is anxiety experienced in social situations (Watson & Friend, 1969). Individuals with social anxiety may experience fear and avoidance of social situations and negative evaluation from others. Studies that have examined the impact of anxiety on performance have suggested that the most detrimental component of anxiety on performance is the cognitive component, such as negative thoughts and rumination (Cassady & Johnson, 2002). Social anxiety can negatively impact performance through increased self-focus (Clark and Wells, 1995; Freda & McManus, 2008) and increased reliance on safety behaviours, such as planning and avoidance (Clark and Wells, 1995; Freda and McManus, 2008), leading to the disruption and overloading of cognitive processes, such as working memory (Ashcraft, 2002; Ashcraft & Krause, 2007; Eysenck & Calvo, 1992). For example, people may devote their attention to rumination on internal thoughts, diverting resources from engaging effectively with the external environment. Research shows that performing safety behaviours and self-focus can interfere with performance in a social situation by diverting attention from the external world (Freda & McManus, 2008), but these behaviours are *believed* to be helpful by those performing them (Freda & McManus, 2008). Following this line of research, social anxiety may impact on monitoring performance in a social domain.

The relationship between social anxiety and social competence is not always clear (Rapee & Heimber, 1997). According to cognitive models of social anxiety (Clark & Wells, 1995; Heimberg, Brozovich, & Rapee, 2010; Hofmann, 2007) distorted beliefs about the world (e.g., unrealistic goals for performance) and the self (e.g., poor social self-efficacy) contribute to emotional dysregulation (e.g., fear) and heightened self-focus (e.g., rumination) in social situations, which can interfere with performance (Freda & McManus, 2008) and further perpetuate social anxiety (Clark, 2001; Clark & Wells, 1995). Rapee & Heimber (1997), proposed that negative beliefs about the world and self may result from actual social skills deficits, negatively distorted perceptions of one's social skills, or both.

Studies that have examined the role of social anxiety in emotional face recognition *skill* have yielded somewhat mixed findings. A negative interpretation bias (e.g., misidentifying neutral expressions as sad) was shown in adults with social anxiety (Arrais, Machado-de-Sousa, Trzesniak, Filho, Ferrari, Ororio et al., 2010; Winton, Clark, Edelmann, 1995) and children with social anxiety (Simonian, Beidal, Turner, Berkes, & Long, 2001). However, one study showed that adults with social anxiety can discriminate between neutral and non-neutral faces, but are relatively less able to discriminate between facial expressions of varying intensities (Montagne, Schutters, Westenbery, van Honk, Kessels, & de Haan, 2006). This effect was similarly shown for children with social anxiety (Wong, Beidel, Sarvor & Sims, 2012). Individuals with social anxiety also show a heightened sensitivity towards threatening facial expressions. With this attentional bias, they are quicker to recognize threatening facial expressions than individuals without social anxiety (Coles & Heimberg, 2005; Mogg, Philippot, & Bradley, 2004; Pishyar, Harris, & Menzies, 2004). While interpretation biases in adults with social anxiety have not been observed consistently across studies (Kolassa, Kolassa, Musial and Miltner, 2007; Kolassa & Miltner, 2006), adults with social anxiety were found to have significant impairments in emotional face recognition in a meta-analysis of the literature (Demenescu, Kortekaas, den Boer, & Aleman., 2010). Thus, higher social anxiety may be associated with lowered success on the emotional face recognition task.

No known studies have examined emotional face recognition *confidence* in relation to social anxiety and research on how anxiety may impact confidence judgments in other domains is sparse. Given cognitive models of social anxiety have proposed that individuals have negative views about their own performance (Hoffman, 2007), higher social anxiety may associate with lower confidence for emotional face recognition. However, given research has also proposed that increased self-focus and safety behaviours in individuals with social anxiety interferes with performance but are *believed* to be helpful by those performing them (Freda and McManus, 2008), higher social anxiety may associate with higher confidence. Studies that have explored the association between anxiety and confidence judgments in academic contexts have provided mixed findings. Some studies have provided support for the relationship between increased anxiety and decreased monitoring accuracy (Everson, Smodlaka, &Tobias, 1999; Legg & Locker, 2009). For example, Everson et al.(2009) demonstrated that increased test-taking anxiety was related to decreased metacognitive monitoring on a reading comprehension task, independent of actual reading ability. In contrast, some research demonstrates that certain levels of state anxiety can be helpful to high-risk situations, given that feelings of uncertainty can lead to the allocation of greater effort and attentional resources (Macher, Papousek, Ruggeri, & Paechter, 2015). Massoni (2014) measured the effect of worry on calibration and discrimination for a perceptual task and demonstrated that higher levels of worry did not impact on actual task performance, but did positively impact on metacognitive monitoring. Some research has found no association between anxiety and confidence judgments (Jiang & Kleitman, 2014). Considering these differing views, it has been proposed that a new area of investigation should explore connections between metacognition and anxiety in an educational domain (Morsanyi, Cheallaigh, and Ackerman, 2019).

Following both cognitive models of social anxiety (Hoffman, 2007) and metacognitive models of monitoring accuracy (Ackerman, 2017), self-reported social anxiety may associate with monitoring accuracy indices for emotional face recognition. If high social anxiety cues lower confidence (Ackerman, 2019) and impairs performance (Demenescu, Kortekaas, den Boer, & Aleman, 2010), higher social anxiety may associate with greater resistance to the overconfidence bias that is typically seen in the general population. However, if high social anxiety does cue lower confidence (Ackerman, 2019), social anxiety may also associate with a diminished ability to discriminate between correct and incorrect choices (i.e., resolution). Therefore, it was hypothesized that increased self-reported social anxiety would correlate with lower confidence judgments and better calibration (i.e., less overconfidence), but lower discrimination between correct and incorrect answers (i.e., resolution) for the emotional face recognition task.

Emotion Differentiation. Differentiation of emotions represents the ability to identify and classify one's emotional experiences into discrete categories (Barrett, Gross, Christensen, & Benvenuto, 2011; Kashdan, Ferssizidis, Collins, & Muraven, 2010). Individuals with a diminished ability to understanding and describe what they are feeling (i.e., low differentiation) have difficulty recognizing detail about an emotional state beyond broadly positive or negative valence. For example, they may use "I feel bad" to describe a negative experience and "I feel good" to describe a positive experience. In contrast, individuals with an ability to understand and classify their emotions, have less extreme attitudes. For example, "I feel somewhat frustrated" may be used to describe a negative experience. Individuals with increased differentiation of emotions may have greater capacity to classify their confidence judgments into discrete categories (e.g., "I am 60 percent confident") than individuals with decreased differentiation of emotions (e.g., "I am not confident"). The latter description is broad and extreme, and may precipitate not only increased intensity of emotions (see Barrett et al., 2011), but also extreme confidence judgments. No known studies have examined differentiation of emotions in relation to monitoring accuracy. However, studies examining alexithymia (a condition characterized by low differentiation of emotions) in relation to social anxiety have reported increased social anxiety with increased alexithymia (Kaur & Kaur, 2015), suggesting that the differentiation of emotions may impact on self-perceptions (e.g., “I am bad at this”) during metacognitive monitoring. Differentiation of emotions was considered in the current study across emotion recognition and general knowledge domains, as detailed awareness of one’s emotional states may impact self-perceptions and metacognitive monitoring regardless of domain-content. It was hypothesized that increased differentiation of emotions would be correlated with better calibration and better discrimination between correct and incorrect answers.

Overall, self-perceptions of empathy, social anxiety, and differentiation of emotions were measured in the current study and explored as correlates of monitoring accuracy for emotional face recognition. It was hypothesized that calibration and resolution for emotional face recognition would be positively correlated with perceived empathic ability and differentiation of emotions. It was hypothesized that self-reported social anxiety would be positively correlated with calibration and negatively correlated with resolution. Relationships between these self-perceptions and monitoring accuracy in a general knowledge domain were also explored for cross-domain comparisons but not expected.

Cognitive Emotion Regulation. Certain thinking dispositions may impact a person’s confidence judgments or the metacognitive cues that underlie the formation of a person’s confidence judgments. For example, in the clinical psychology literature, the tendency to think in black and white or overgeneralized terms can influence perceptions about the self and others and contribute to anxiety (i.e., metacognitive cue) (Greene, 1999; Stallard, 2005; Weems, Berman, Silverman, & Saavedra, 2001; Epkins, 1996). Emotion regulation refers to control or influence over one's emotional experiences (Gross, 1999). Cognitive emotion regulation involves using various cognitive regulatory strategies such as rationalization or reappraisal to decrease, maintain, or increase negative and positive emotions (Masters, 1991; Langston, 1994). This includes influencing or controlling which and how strongly emotions are experienced (Gross, 1999; Gross & Thompson, 2007; Calkins & Hill, 2007) which may alleviate stress on cognitive processes during decision-making (Cassady & Johnson, 2002). In addition to alleviating demands on cognitive processing to facilitate optimal decision-making, emotion regulation may also facilitate unbiased assessment of self and others (e.g., Nelson and Perry, 2015; Lemerise & Arsenio, 2000). Individuals enter a social situation with emotional predispositions or vulnerabilities that influence the intensity of their reactions to social cues (Eisenberg & Fabes, 1992; Rothbart & Derryberry, 1981) and individuals differ in how well they can regulate or control their emotional reactions (Eisenberg et al., 1997). For example, a highly sensitive individual may be more likely to experience anxiety in response to a social cue, and if that individual is not able to effectively regulate that emotion, then it might interfere with their ability to correctly monitor their understanding of the situation. An important consideration is whether poor emotion regulation impacts monitoring accuracy for emotional face recognition. Therefore, cognitive emotion regulation was considered in the current study.

It was hypothesized that increased cognitive emotion regulation would be correlated with increased calibration and resolution abilities in an emotional face recognition domain. Given cognitive emotion regulation may impact on monitoring accuracy regardless of domain-content, associations between this scale and monitoring accuracy in a general knowledge domain was also explored but not expected.

Behavioural Correlate

A final consideration would be to gain an understanding of how metacognitive skills in a social domain may relate to social behaviours. Overconfidence has been associated with behaviour in real-world contexts, such as various risk-taking behaviors, externalizing behaviors, and substance use (Parker & Fischhoff, 2005). In addition, educational research has demonstrated that the use of metacognitive skills are important for guiding a students' learning in various academic areas, (e.g., mathematics; Pugalee, 2001; reading; Presley, 2002; and writing; Pugalee, 2001). Following this line of research, a final, exploratory goal in the current study was to examine potential behavioural correlates of metacognitive monitoring abilities in a social domain. Specifically, confidence judgments for emotional face recognition were further explored in relation to self-reported social-emotional challenges. Participants were asked to provide a self-reported rating of emotional, behavioural, or social challenges. It was hypothesized that overconfidence and poor discrimination on the emotional face recognition task would correlate with and predict self-reported social-emotional challenges in university students.

Study Overview

The goals of the current study were to build and extend research on monitoring accuracy in a social domain, which has been generally understudied in both the metacognitive and cognitive science literature (see Figure 1 for an overview). First, the question of domain-generality versus domain-specificity was explored through the use of tasks that measured emotional face recognition and general knowledge. Calibration and resolution were measured and compared between domains. Second, calibration and resolution, two related yet separate indices of monitoring accuracy, were compared within domains. Third, relationships between cognitive abilities and monitoring accuracy were assessed. Fourth, relationships between aggregate task-specific judgments (i.e., predictive and postdictive confidence; difficulty; effort) were explored as correlates of monitoring accuracy across domains. Fifth, additional person-related variables (i.e., self-perceptions and social-emotional dispositions) were explored in relation to monitoring in an emotional face recognition domain. Specifically, relationships between monitoring accuracy and self-report measures of perceived empathic ability, social anxiety, and emotion differentiation and regulation were explored. Last, the potential relationship between monitoring accuracy in an emotional face recognition domain and a behavioural correlate was explored (i.e., social, emotional, or behavioural challenges). Research hypotheses were as follows:

Summary of Hypotheses

Hypothesis 1. As has been reported in the literature, overconfidence was expected across both tasks (Bjork et al., 2013; Soderstrom et al., 2015; Stanovich et al., 2016). Given that a Bias Index score greater than "0" indicates overconfidence, Bias was expected to be greater than "0" on both tasks. As has been reported in the literature, resolution was expected across both tasks (e.g., Kelly and Metcalfe, 2011). Given that a gamma value of "0" indicates a lack of resolution, resolution was expected to be greater than zero across all tasks. Resolution and calibration were not expected to be correlated within experimental tasks.

Hypothesis 2. Better calibration on the emotional face recognition task was expected to be correlated with better calibration on the general knowledge task (Jackson & Kleitman, 2014; Schraw et al., 1995).

Hypothesis 3. Resolution, although less explored across domains, has been found to be less consistent (Jackson et al., 2016). It was hypothesized that resolution scores would not be correlated across tasks, especially given the differences in domain content.

Hypothesis 4. Based on previous research (Bruine de Bruin et al., 2007; Stanovich & West, 1998; Stanovich et al., 2016), working memory and intellectual abilities were expected to be positively correlated with higher calibration and resolution across both general knowledge and emotional face recognition tasks.

Hypothesis 5. Calibration and resolution scores were expected to be positively correlated with predictive and postdictive confidence and negatively correlated with task difficulty and effort ratings (Ackerman, 2019; Chen, 2002; Hsu et al., 2018).

Hypothesis 6. Confidence for emotional face recognition was expected to be positively correlated with perceived empathetic ability and negatively correlated with social anxiety, as was similarly suggested for self-perceptions in academic domains (Ackerman, 2019). Resolution and calibration for emotional face recognition was expected to be positively correlated with perceived empathic ability, differentiation of emotions, and cognitive emotion regulation. Social anxiety was expected to be negatively correlated with resolution and positively correlated with calibration.

Hypothesis 7. Correlates of monitoring accuracy, including cognitive abilities and task-specific judgments, were expected to predict calibration and resolution scores for both emotional face recognition and general knowledge. In addition, self-perceptions and cognitive emotion regulation were expected to predict calibration and resolution for emotional face recognition.

Hypothesis 8. Indices of monitoring accuracy for emotional face recognition were expected to correlate with and predict self-reported social-emotional challenges.

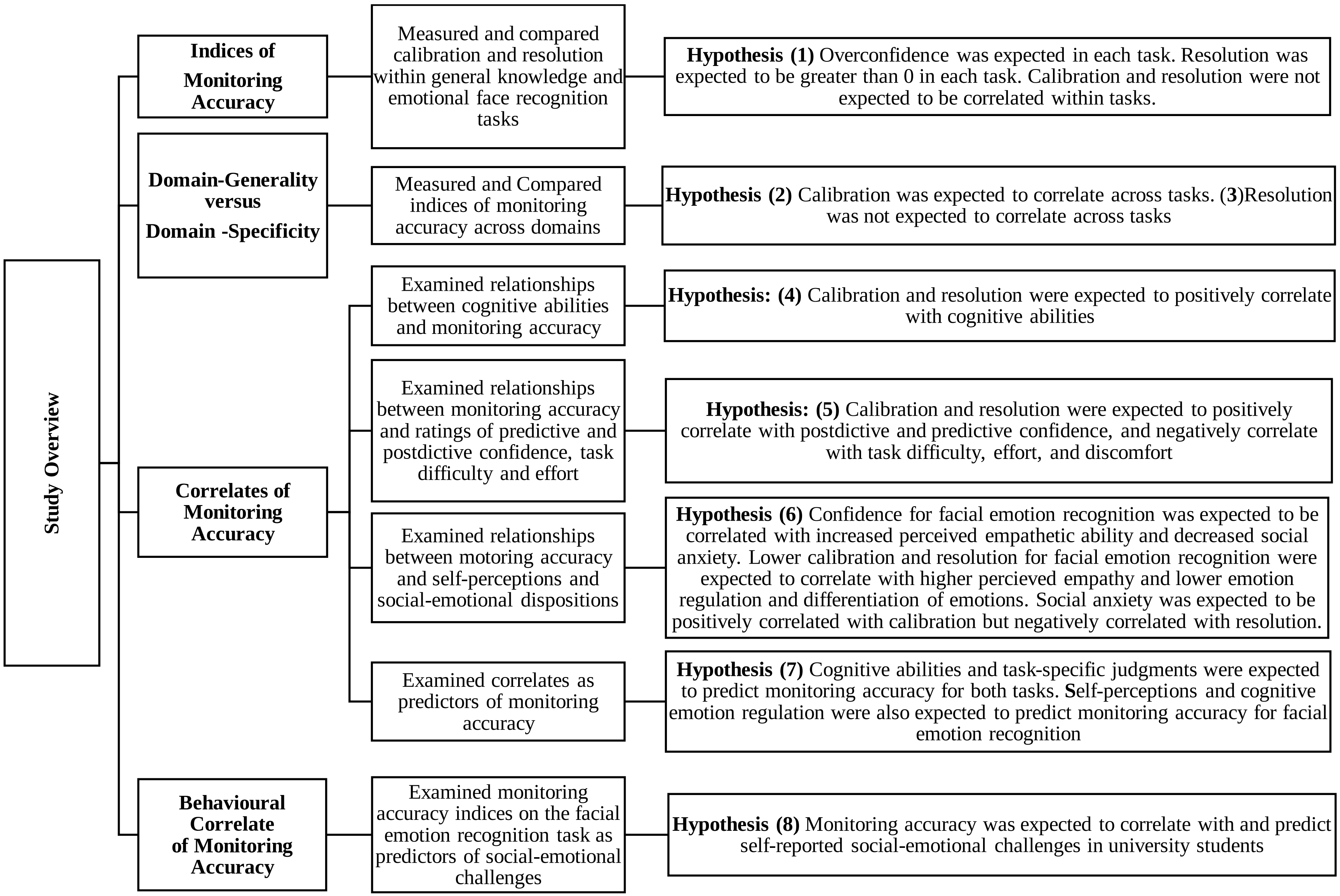


Figure 1. Study overview. This figure outlines the main study objectives, research questions, and hypotheses.

Methods

Participants

Participants consisted of 136 undergraduate students attending York University in Toronto, Ontario (*M* age = 20.43 years, *SD* = 2.77; age range 18 – 30 years of age; 34 males and 102 females). Participants were recruited from an undergraduate participant pool. The data were collected during a single term. Participants received two course credits for participating in this study.

Initially, 153 participants took part in the study. However, 17 participants were excluded from data analyses (four participants reported English as a second language and identified language-related difficulties in understanding and/or completing study tasks; 11 participants did not complete full task battery due to time constraints; one participant did not complete the tasks correctly; and one participant did not have adequate variability to calculate a resolution score). Therefore, the final sample included in the data analyses was 136 participants. Most participants were in their first year of the undergraduate program (*n* = 77, 56.5%), 28 were in their second year (20.6%), 18 were in their third year (13.2%), eight were in their fourth year (5.9%), and five were post fourth year students (3.7%). In terms of ethnicity, 43 participants reported White/European (31.6%), 33 reported South-Asian (24.3%), 23 reported Asian (16.9%), 12 reported Black (8.8%), five reported Arab (3.7%), six reported Latino-Hispanic (4.4%) and 14 reported Other (10.3%). Twenty-seven participants reported that their current academic average was 80-100% (19.9%), 62 participants reported that their current academic average was 70-79% (45.6%), 36 participants reported that their current academic average was 60-69% (26.5%), five reported that their current academic average was 50-59% (3.7%) and six participants did not provide a response (4.4%). Table 2 includes detailed description of participant demographic information, including gender, ethnicity, socio-economic status, year of study and current academic average. Appendix A contains the demographic measure.

Table 2
Descriptive Statistics for Socio-Demographic Variables (N=136)

Socio-Demographic Variable	Frequency	Percent
Gender		
Male	34	25.0
Female	102	75.0
Year in university		
1st year undergrad	77	56.5
2nd year undergrad	28	20.6
3rd year undergrad	18	13.2
4th year undergrad	8	5.9
Post 4th year undergrad	5	3.7
Ethnicity		
White/European	43	31.6
Black	12	8.8
Asian	23	16.9
South-Asian	33	24.3
Arab	5	3.7
Latino-Hispanic	6	4.4
Other	14	10.3
Mother highest level of education		
Partial high school	17	12.5
High school graduate	22	16.2
Partial college/university	26	19.1
College/university education	48	35.3
Graduate/Professional degree	20	14.7
No response	3	2.2
Father highest level of education		
Partial high school	13	9.6
High school graduate	23	16.9
Partial college/university	12	8.8
College/university education	49	36.0
Graduate/Professional degree	36	26.5
No response	3	2.2

Family income		
Well below or below average	18	13.2
Average	75	55.1
Well above or above average	39	28.7
No response	4	2.9
Current academic average		
50-59%	5	3.7
60-69%	36	26.5
70-79%	62	45.6
80-100%	27	19.9
Self-reported socio-emotional difficulties		
No	74	55.06
Yes – minor difficulties	43	32.57
Yes – definite difficulties	14	10.61
Yes – severe difficulties	1	0.75

Measures

Monitoring Accuracy Tasks. Two experimental tasks were developed to measure monitoring accuracy. A general knowledge task was developed to measure monitoring accuracy in a general knowledge domain and an emotional face recognition task was developed to measure monitoring accuracy in a social information processing domain. Both tasks had the same confidence rating scale and the same multiple-choice format with four choices. The tasks were developed to be matched on total number of items, accuracy rates, and confidence ratings. These tasks were piloted using a paper-and-pencil version of the tasks with a sample of 18 undergraduate and graduate students to ensure that the instructions were clear and that items were not too easy or too difficult. Overall success rate on the pilot tasks was greatest for the emotion recognition task (69 %) followed by the General Knowledge Task (60 %). Items were judged to be at a moderate level of difficulty, which permitted adequate variability in performance to calculate the different monitoring accuracy indices. Appendix B contains the pilot data for each task including item-by-item success rate.

Confidence rating scale. Confidence judgments were measured on an item-by-item basis with each item on the general knowledge and emotional face recognition tasks being followed with a confidence rating scale. The general knowledge and emotional face recognition tasks contained four multiple choice item responses. Therefore, confidence judgments were marked on a scale ranging from 25% (i.e., chance of responding correctly by guessing) to 100% (i.e., absolute confidence in the answer provided). This confidence scale was based on previous confidence rating scales used in the decision-making literature (Bruine de Bruin et al., 2007). Instructions on how to use the confidence rating scale was provided to participants both orally and visually, with examiners providing these instructions on projected PowerPoint slides. Practice trials were also presented before each task, in order to provide feedback and to ensure that participants understood how to use the confidence rating scales.

The general knowledge task was presented first with elaborated instructions for explaining the responses required and examples for completing the confidence ratings. The instructions were as follows:

> In this survey, you will be asked to answer general knowledge questions. There will be four options for each question (A, B, C, and D). Please choose the option that you think is correct for each question.
>
> Following each question, you will be asked to indicate how sure you are about your answer on a scale from 25% (*Just Guessing, Not Confident at All)* to 100% (*Very Confident*). Since there are 4 options for each question, 25% would mean you are just guessing and you are not very confident. Then, 100% would mean you are very confident and certain. Respond to each confidence rating immediately after answering the question. Spend no more than 5 seconds on each confidence rating. The following is an example:
>
> **What is the name of a young sheep?**

A. Lamb
B. Calf
C. Baby
D. Steer

| . . . . |. . . . |. . . . |. . . . |. . . . |. . . . |. . . . |. . . . |. . . . |. . . . |. . . . |. . . . |. . . . |. . . . |. . . . |. . . . |

25% 35% 45% 55% 65% 75% 85% 95% 100%

Not Confident At All	Somewhat Confident	Pretty Confident	Very Confident

Please write the number that you just circled: ______________________

Participants were instructed to first select one of the four choices (A, B, C, or D) and then to indicate how confident they were in their choice by circling the number on the confidence scale and writing down the number that they circled. Following these instructions, participants were provided with examples of three possible responses. In this sample item, option A (i.e., Lamb) was the correct response. Using the choice of "Lamb" (i.e., Choice A) for this example, participants were instructed that a response of 25% would indicate that they are guessing, a response of 50% would indicate that they are somewhat confident, and a response of 100% would indicate that they are 100% sure. This example was also displayed visually on the response form.

General Knowledge. The general knowledge task assessed knowledge about a broad range of facts and topics. The items were selected from published general knowledge norms reported in Tauber, Dunlosky, Rawson, Rhodes, and Sitzman (2013; which were updated from norms first published by Nelson & Narens, 1980). These norms provided recall accuracy and confidence ratings. Therefore, items were selected to reflect a wide range of difficulty and confidence ratings, while avoiding floor and ceiling effects. Each question was presented separately, directly followed by the confidence rating scale. The following item is an example:

What is the name of the largest ocean on earth?

A. North Sea
B. Atlantic
C. Pacific
D. Indian

How confident are you that your response is correct? Please circle the number that best represents how sure you are about the answer you just provided.

| . . . . | . . . . | . . . . | . . . . | . . . . | . . . . | . . . . | . . . . | . . . . | . . . . | . . . . | . . . . | . . . . | . . . . | . . . . |

25% 35% 45% 55% 65% 75% 85% 95% 100%

Not Confident At All — Somewhat Confident — Pretty Confident — Very Confident

Please write the number that you just circled: ______________________

Option C was the correct response for this item. Initially, the general knowledge task consisted of 24 items and participants completed all 24 items. However, to ensure adequate variability in confidence ratings to calculate the calibration and resolution indices, items with overall accuracy corresponding to 10% or less, as well as items with overall accuracy corresponding to 90% or more were subsequently removed (for accuracy variability, see Table 3). Therefore, eight items (three below 10% accuracy and five above 90% accuracy) were removed, leaving 16 items on this task for analysis[1]. General knowledge items have shown high reliability in previous studies (e.g., Mansueti, Frias, Bub, and Dixon, 2009). Cronbach's alpha

[1] For the general knowledge and emotion recognition tasks, due to a technical printing error, 38 participants missed completing two items on each of these tasks. We compared overall accuracy between those participants who missed the two items and the remaining 98 participants. There were no differences in overall accuracy. Thus, scores of these 38 participants were pro-rated for the statistical analyses.

for the general knowledge task in the current study was 0.36. Appendix C contains test question, item-by-item success rate, and removed items.

Table 3
Accuracy Variability for the General Knowledge Task (N = 24 items)

Percent Accuracy	Frequency	Percent
0-10	3	12.50
10-20	2	8.33
20-30	1	4.17
30-40	2	8.33
40-50	4	16.67
50-60	4	16.67
60-70	1	4.17
70-80	1	4.17
80-90	2	8.33
90-100	5	20.83

Note: items below 10% accuracy and above 90% accuracy were removed from analyses. Thus, eight items were removed from analyses.

Emotional Face Recognition Task. The items for this task were taken from the Reading The Mind in the Eyes task developed by Baron-Cohen (2001; abbreviated to the eyes task by authors). The eyes task is a popular measure of mental state recognition. This task requires participants to identify another's affective state based on a cropped image of the eyes. Facial expressions reflecting positive (friendly, confident, playful), negative (regretful, hostile, panicked) and neutral (thoughtful, regretful, insisting) affective states were selected. Baron-Cohen (2001) published the proportion of participants (comprised of a group of undergraduate students, mean age 20.8, and adults, mean age 28.0) who answered items correctly on the Reading the Mind in the Eyes task. Therefore, items were selected to reflect a wide range of difficulty, across the positive, negative and neutral affective state groups. Each item in this task was presented on a separate page followed by the confidence rating scale.

The following item is an example:

A. *Apologetic*
B. *Friendly*
C. *Uneasy*
D. *Dispirited*

How confident are you that your response is correct? Please circle the number that best represents how sure you are about the answer you just provided.

| . . . . |. . . . |. . . . |. . . . |. . . . |. . . . |. . . . |. . . . |. . . . |. . . . |. . . . |. . . . |. . . . |. . . . |. . . . |

25% 35% 45% 55% 65% 75% 85% 95% 100%

Not Confident At All — Somewhat Confident — Pretty Confident — Very Confident

Please write the number that you just circled: ______________________

Option C (Uneasy) was the correct response for this item. To help ensure that comprehension problems with the words do not interfere with performance, a glossary of all the emotional state terms included in this task was provided and participants were encouraged to refer to this glossary. Appendix D contains the glossary.

Initially, the emotional face recognition task included 24 items and participants completed all 24 items. However, to ensure adequate variability in confidence ratings to calculate the calibration and resolution indices, items with accuracy corresponding to 10% or less, as well as items with accuracy corresponding to 90% or more were subsequently removed from analyses (for accuracy variability, see Table 4). Therefore, one item was removed, leaving 23 items for analyses.

Table 4
Accuracy Variability for the Emotion Recognition Task (N = 24 items)

Percent Accuracy	Frequency	Percent
0-10	0	0
10-20	0	0
20-30	0	0
30-40	0	0
40-50	1	4.17
50-60	9	37.5
60-70	6	25
70-80	6	25
80-90	1	4.17
90-100	1	4.17

Note: items below 10% accuracy and above 90% accuracy were removed from analyses. Thus, one item was removed.

Previous research has shown that the eyes task has acceptable test-retest reliability (Hallerbäck et al., 2009; Vellante et al., 2012; Prevost et al., 2014; Khorashad et al., 2015). Some previous studies have shown that eye task scores have sufficient internal consistency (Dehning et al., 2012; Prevost, Carrier, Chowne et al., 2013) while other studies have shown that eye task scores have low internal consistency (Harkness, Jacobson, Duong, & Sabbagh, 2010; Vellante et al., 2012; Voracek & Dressler, 2006). Cronbach's alpha for the emotional face recognition task in the current study was 0.34. Previous research has shown that Eye task scores show convergent validity evidence with Theory of Mind measures (e.g., Strange Stories Test; Torralva et al., 2009; Ferguson and Austin, 2010; Kirkland et al., 2012a). The emotional face recognition task in the current study showed convergent validity with the Toronto Empathy Questionnaire (this measure was an assessment of self-reported empathic ability and is described below).

Monitoring Accuracy Indices. The degree of fit between subjective confidence judgments and objective performance can be examined in various ways. The present study focused on two types of indices: (1) calibration and (2) resolution. Figure 2 summarizes indices used in this study.

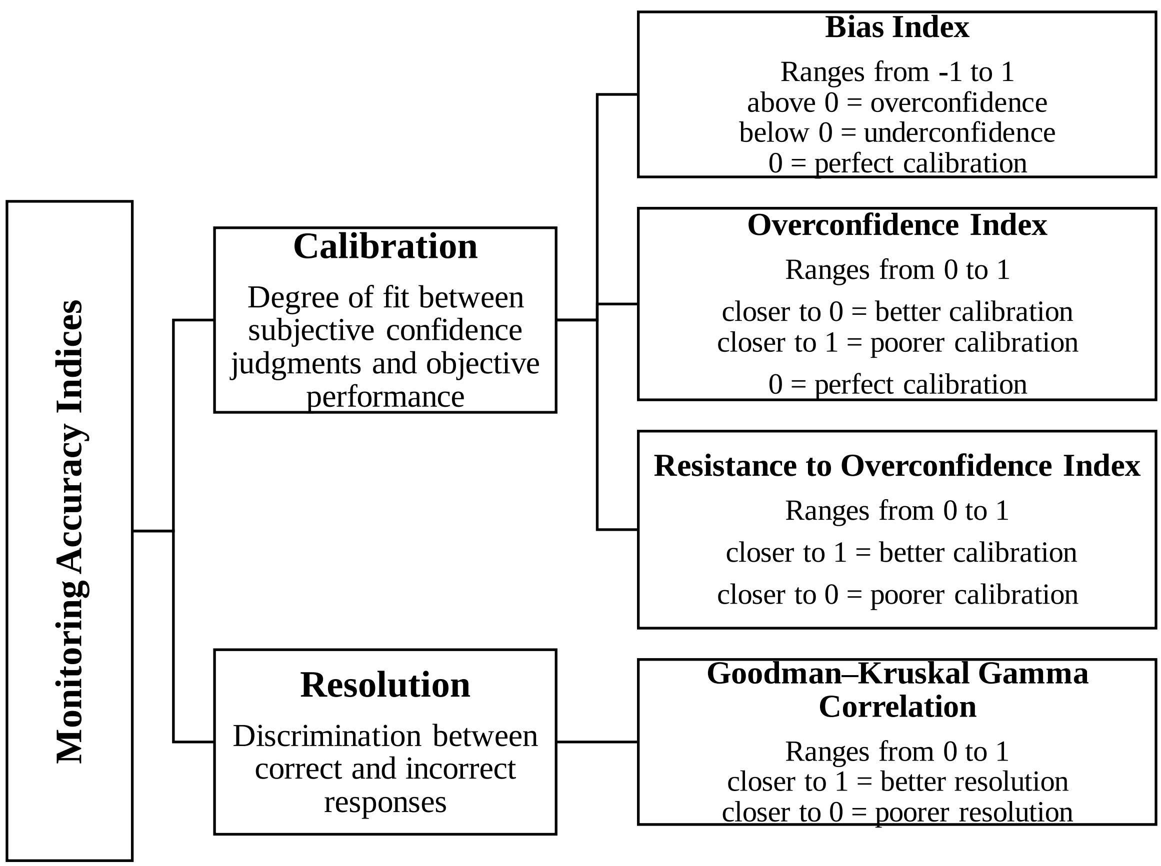


Figure 2. Indices of Monitoring Accuracy.

Calibration. Calibration broadly refers to how precisely one's confidence matches one's performance on a test. The present study focused on three calibration indices: (1) the Bias Index (2) the Overconfidence index and (3) the Resistance to Overconfidence Index.

Bias Index (Ronis & Yates, 1987; Yates, Zhu, Ronis & Wang, 1989). The Bias Index captures the degree of overconfidence or underconfidence on a task (Was, 2014). It is the difference between the mean subjective probability of a correct answer (i.e., mean confidence rating) and the proportion of correct answers (i.e., percent of items answered correctly). The Bias Index is represented by the following equation:

Bias Index = (mean confidence – percent correct)

Scored differences can be positive or negative. Bias Index scores range from -1 to 1, where positive and above zero scores indicate overconfidence, and negative and below zero scores indicate under-confidence. Scores close to zero indicate low bias, with confidence being matched to performance.

Overconfidence Index (Bruine de Bruin; Parker & Fischhoff, 2007). For investigations including correlations, where scores are unidirectional and range from zero to one, the Bias Index cannot be examined in correlational analyses, as both ends of the Bias Index measure a type of miscalibration (see Figure 3).

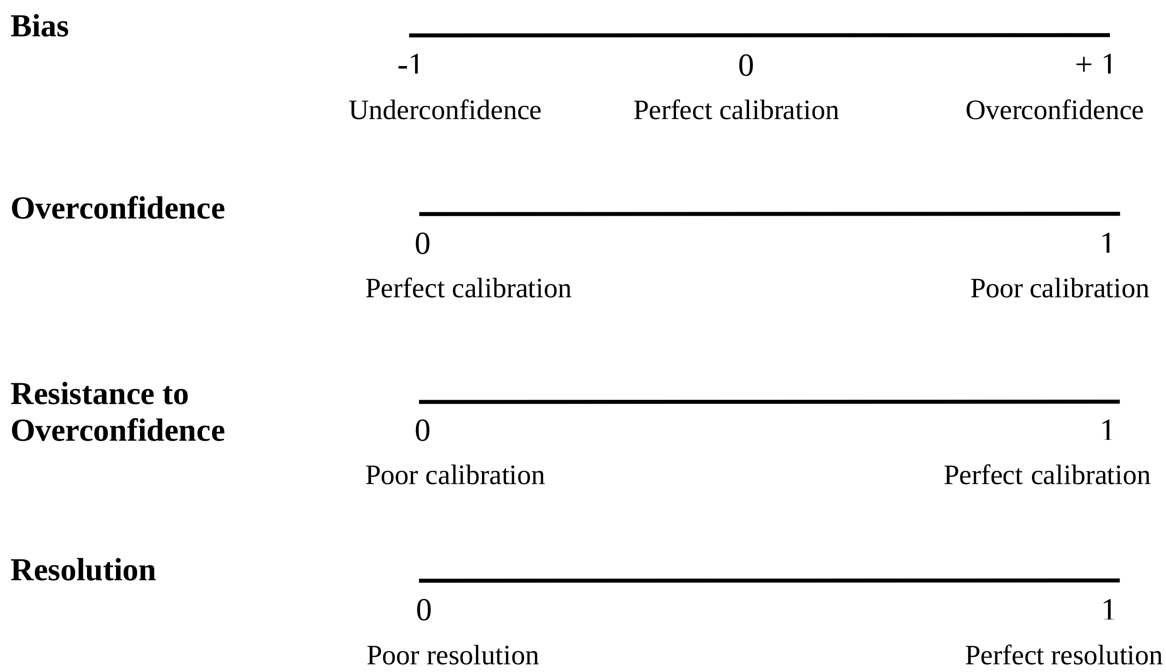


Figure 3. *Value ranges for calibration and resolution indices.*

Therefore, a measure to index the magnitude of the Bias score was needed. As the tendency for participants is to display overconfidence rather than underconfidence, and to be consistent with the decision-making literature (Stanovich, West & Toplak, 2016), we called this index the Overconfidence Index. The Overconfidence Index provided a measure of the difference between one's confidence and one's performance on a given test. It is the absolute difference between mean confidence ratings and percentage of items correct. The Overconfidence Index is represented by the following equation:

$$\text{Overconfidence Index} = |\ \text{mean confidence} - \text{percent correct}\ |$$

Values range from 0-1, with higher scores reflecting a greater difference between confidence and performance (i.e., poorer calibration) and lower scores reflecting a smaller difference between confidence and calibration (i.e., better calibration). A score of zero means that confidence is perfectly matched with performance.

The Resistance to Overconfidence Index (Bruine de Bruin; Parker & Fischhoff, 2007*).* This calibration index was formulated for the current study and is the inverse of the Overconfidence Index described above. It is represented by the following equation.

Resistance to Overconfidence Index = 1 - | mean confidence - percent correct |

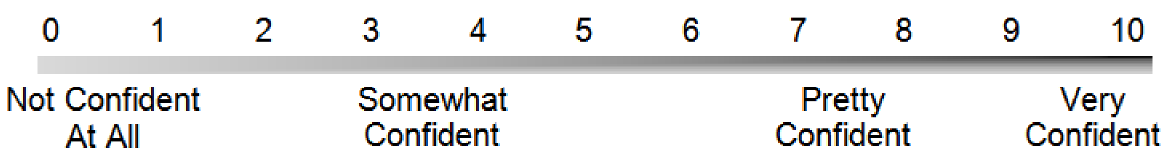
or 1 – Overconfidence Index

As described above, the Overconfidence Index values range from zero to one, with higher scores reflecting poor calibration. The Resistance to Overconfidence Index scores also range from zero – one, but higher scores reflect better calibration. A score of one means that confidence is perfectly matched with performance. The Resistance to Overconfidence Index provided a score that had the same direction so that a higher score reflected better monitoring accuracy, consistent with the resolution index.

Resolution (Ackerman & Goldsmith, 2011). Resolution reflects the extent to which and individual can distinguish between an event's occurrence (i.e., a correct answer) and non-occurrence (i.e., an incorrect answer). High resolution occurs if participants can accurately assign higher probability of being correct to correct answers (e.g., 100% confidence in a correct answer) and a lower probability of being correct to incorrect answers (e.g., 25% confidence in an incorrect answer). The Resolution Index can be measured with the use of the Goodman–Kruskal Gamma correlation and involves a within-subjects measure of the relationship between confidence judgments and the correctness of each individual item. Resolution values initially ranged from -1 to 1, with positive and higher values representing higher resolution (e.g., 100% confidence for correct answers and 50% confidence for incorrect answers) and negative and lower values representing lower resolution (e.g., 100% confidence for correct answers and 90% confidence for incorrect answers). In the current sample, 11 participants had resolution values below one. These values were replaced with scores of zero, which did not impact tests of significance. Thus, resolution values ranged from zero – one, with higher values still representing better resolution and lower values still representing poorer resolution.

Task-Specific Judgments.

Predictive Confidence. Prior to completing each task, participants were asked for an aggregated predictive confidence judgment regarding their ability to answer general knowledge questions (i.e., How confident are you in your ability to correctly answer general knowledge questions, such as “What is the capital of Greece?”), and their ability to read the emotional state of a person (based on only an image of the person’s eyes). Participants were asked to rate their confidence judgments on the following scale:



Postdictive judgments. Following each of experimental task, participants were asked to rate their perceived performance again on the same scale described above (e.g., “How confident are you in your ability to correctly solve general knowledge questions?”; 0 = Not Confident at All; 10 = Very Confident).

Task difficulty, effort, and effort required. After each experimental task, participants were asked to rate their aggregated judgments of perceived difficulty (i.e., “How difficult did you find this task?”; 0 = Not Difficult at All; 10 = Very Difficult), required effort (i.e., “How much effort was required of you to complete this task?”; 0 = No Effort at All; 10 = Extreme

Effort) and the feeling associated with expending effort (i.e., "How did using that level of effort make you feel?"; 0 = Extremely Pleasant; 10 = Extremely Unpleasant). Tables 5, 6, and 7 contains means, standard deviations, ranges, and normality estimates for the aggregated ratings of each experimental task.

The task-specific judgments ratings were normally distributed, with skewness values below three and kurtosis values below 10 (Kline, 2005). Based on the cue-utilization theory of metacognitive reasoning (Ackerman and Thompson, 2017; Ackerman, 2019), individuals may use aggregate judgments of predictive and postdictive performance, task difficulty, and their experience of effort when forming a judgment about their performance on a task. In order to explore individual differences in monitoring accuracy, it was important to measure these aggregate ratings in addition to item-by-item confidence (see Ackerman & Thompson, 2017).

Cognitive Ability. Cognitive ability, as assessed in the present study, encompassed two domains of cognitive functioning: (1) verbal and nonverbal ability and (2) working memory. The Shipley-2 was used to assess reasoning ability and the Reading Span Task was used to assess working memory.

Intellectual Ability. The Shipley-2 (Shipley, Gruber, Martin, & Klein, 2009***)*** provided an estimate of general intelligence. It included two subtests: Vocabulary and Block Patterns. The Shipley-2 is a standardized test that can be administered individually or in group format, with an administration length of 20 minutes. On the Vocabulary subtest, participants were asked to choose from four alternatives the definition that most closely matches the target word. Scores on the Vocabulary subtest are thought to reflect crystalized abilities. On the Block Patterns subtest, participants were asked to choose from four alternatives the block that best completes the specified design. Scores on the Block Patterns subtest are thought to represent fluid reasoning abilities. Raw scores (not age corrected) were standardized and summed to create a composite score of general intelligence, which we called the Intelligence Score Composite. The Shipley Vocabulary test has been reported to range from .85 to .92 across age groups and the Block Patterns has been reported to range from .88 to .94 across age groups for internal consistency (Shipley et al., 2009). The Shipley-2 shows convergent validity with conventional intelligence measures, such as the Weschler scales (i.e., Weschler Adult Intelligence Scale- Third Edition; Weschler Intelligence Scale for Children-Fourth Edition; Shipley, Gruber, Martin, & Klein, 2009). Higher scores indicated better intellectual abilities. Scores were normally distributed.

Working Memory. The reading-span task, based on the methods of Turner and Engle (1989), provided a measure of working memory. This task was group administered with sentences presented on PowerPoint slides on a projector screen. This task included 12 blocks, each block consisting of two, three, four, or five sentences. There were a total of 42 items (two sets of two, three, four and five sentences). Participants were provided with a response form to follow along with while direct instructions were given by the examiner. Participants were provided with the following instructions:

> You will see a sentence on the screen. Your job is to read the sentence out loud, along with me. As soon as you have finished reading the sentence, decide if the sentence is True or False by checking off either True or False on the sheet of paper in front of you.
> There will be 12 sets of sentences, each set containing 2-5 sentences. After each set, you will be prompted to write down the last word of each sentence from that set (e.g., "What was the last word in each sentence that you read in Set #1?) Don't worry about spelling!
> **Please put your pencils down as soon as you have finished writing the words.**

Participants were presented with a practice block containing two sentences before the actual test blocks were provided. For each trial, participants were asked to circle on the response form whether sentences were true or false (e.g., "Cucumbers are green"), while also having to commit the last word in each sentence to memory (e.g., "green"). At the end of each block, participants were asked to recall the to-be-remembered words from the entire block. In order to pace participants on the task, participants were asked to read projected sentences aloud along with the examiner. Cronbach's alpha was 0.66 on this task. Scores were normally distributed. Higher scores on this task indicated higher working memory abilities. Appendix D contains the task items.

Self-Perceptions. Self-perceptions measured in the current study included perceived empathic ability, differentiation of emotions, and social anxiety. Participants were asked to rate their level of agreement with various statements on a Likert-Type scale: (1= "disagree strongly", 2 = "disagree moderately", 3= "disagree slightly", 4= "agree slightly", 5= "agree moderately", 6 = "agree strongly"). All variables were normally distributed with skewness values below three and kurtosis values below 10 (Kline, 2005). Appendix D contains the self-perception scales.

Toronto Empathy Questionnaire* (Spreng, McKinnon, Mar, & Levine, 2009*). The Toronto Empathy Questionnaire (TEQ) is a 16-item self-report measure that assessed perceived empathic ability. Example items are "I find that I am 'in tune' with other peoples' moods" and "I become irritated when someone cries" (reverse scored). Scores were summed to create a composite score for perceived empathy. Higher scores represented higher perceived empathy. Research has shown that the TEQ has high reliability and validity (Spreng et al., 2009). The Cronbach's alpha in the current study was .82.

Differentiation of Emotions Scale* (Stanovich, West & Toplak, 2016*). The Differentiation of Emotions Scale (Stanovich, West & Toplak, 2016) was a 14-item self-report measure that assessed insight into one's emotional experiences. Example items are "I am good at describing my feelings to other people" and "sometimes I just don't know how I feel about people" (reverse scored). Scores were summed to create a composite score for the differentiation of emotions. Higher scores were indicative of greater emotional awareness. Research shows that this scale has high reliability and validity (Stanovich, West & Toplak, 2016). The Cronbach's alpha in the current study was .79. The Differentiation of Emotions Scale correlated with Cognitive Emotion Regulation-Questionnaire composite scores (this is a measure of emotion regulation and is described below), offering support for scale validity.

Social Anxiety in University Students (*Bhamani & Hussain, 2012*). The Social Anxiety in University Students scale is a 15-item self-report measure that assessed social anxiety. This measure assessed several facets of social anxiety experienced by students in a university setting including perceived self-image, perceived social image, and perceived peer response of self-image. Example items are "I feel comfortable being introduced to new people" and "I feel comfortable to sit next to opposite gender class fellows". This scale has been shown to have high validity and reliability (Bhamani & Hussain, 2012).

Scores were summed to create a composite score. Initially, higher scores were indicative of less social anxiety. Scores were inversed such that higher scores were indicative of greater social anxiety. There was one extremely low score. Analyses were conducted with and without the extreme score and tests of significance were not impacted. Thus, the score was retained as there was no justification for its removal. The Cronbach's alpha in the current study was = .77.

Cognitive Emotion Regulation. The Cognitive Emotion Regulation Questionnaire: Short (CERQ-S; Garnefski, Kraaij & Spinhoven, 2001) is an 18-item self-report measure that assessed the use of 9 cognitive strategies for regulating one's emotions: Acceptance (e.g., "I think that I have to accept that this has happened"), Rumination (e.g., "I often think about how I feel about what I have experienced"), Positive Refocusing (e.g., I think of pleasant things that have nothing to do with it"), Refocusing on Planning ("I think about how to change the situation"), Positive Reappraisal (e.g., "I think I can learn something from the situation"), Putting into Perspective (e.g., "I tell myself that there are worse things in life"), Catastrophizing (e.g., "I keep thinking about how terrible it is what I have experienced"), Self-Blame (e.g., "I feel that I am the one who is responsible for what has happened"), and Other-blame (e.g., "I feel that others are responsible for what has happened"). According to Garnefski et al. (2001), theoretically more adaptive coping strategies include Acceptance, Refocus on Planning, Positive Reappraisal and Putting into Perspective and theoretically less adaptive coping strategies include Catastrophizing, Rumination, Self-blame, and Other-blame.

An emotion regulation composite score was created for the current study. This composite was comprised of the Positive Reappraisal, Refocus on Planning, Positive Refocusing, Reduced Self-blame, reduced Other-blame, and reduced Catastrophizing subscales. These subscales were included because they have consistently been shown to associate with emotion regulation challenges (i.e., lower anxiety, depression, aggression or relational stress; Martin and Doehn, 2005; Garnefski &Kraaji, 2006; Schroevers et al. 2007; Ramos-Cejudo et al., 2017). The Putting into Perspective and Acceptance subscales were excluded from this composite because research has shown these scales to have questionable validity (Schroevers et al. 2007; Martin & Doehn, 2005). The Rumination subscale was also excluded due to low internal consistency across items comprising this subscale. Higher emotion regulation composite scores indicated higher use of adaptive cognitive coping skills.

Cronbach's alphas ranged from .68 to .90 for all CERQ-S subscales in previous research (Garnefski and Kraaji; 2006). Present Cronbach's alphas ranged from .70 to .90 for CERQ-S subscales. In the current study, the Cronbach's alpha for the CERQ-S composite was .70. The CERQ-S composite significantly correlated with Social Anxiety in University Students

Questionnaire scores and self-reported social-emotional challenges in the present study, offering evidence for construct validity.

Social-Emotional Challenges. Participants were asked to rate their perceived level of difficulties with social, emotional, or behavioural challenges by rating the question "Overall, do you think that you have difficulties in one or more of the following areas: emotions, behaviour, or being able to get along with other people?" on a four-point Likert-type scale (1 = "no"; 2 = "minor difficulties"; 3 = "definite difficulties"; 4 = "severe difficulties"). Higher scores indicated more severe perceived social-emotional challenges. Responses were normally distributed.

Procedure

The current study was part of a larger study investigating confidence judgments in emerging adults. Testing sessions were conducted in group format and were 120 minutes in length. A maximum of 10 participants were tested at one time and each testing session included two examiners to ensure correct and fluid administration. Participants first completed the informed consent and demographic forms. After these forms were collected, participants completed the group-administered working memory task, followed by the Shipley-2 measures. Participants were engaged in the group-administered working memory task presented on PowerPoint slides via the projector screen. Participants completed the reasoning ability tasks individually on paper protocols. Instructions for these tasks were provided aurally and via a projector screen. Following group-administered tasks, participants were asked to independently complete the monitoring accuracy tasks. Instructions for these tasks were provided aurally and via a projector screen as well as in written format. After all experimental tasks were completed, participants were asked to complete a dispositions questionnaire including measures of self-reported empathy, differentiation of emotions, cognitive emotion regulation, and social-anxiety. Tasks from the current study were counterbalanced with other tasks from the larger study, so that participants received the tasks in one of two possible order formats. There were no significant differences in accuracy between the tasks based on presentation order.

Results

Data Management

All variables of interest were first examined for missing values and accurate data entry. Due to a technical printing error, two items on the general knowledge and emotional face recognition tasks were missed by 38 participants. Overall accuracy rates between participants who missed the two items and the remaining 98 participants was compared. There were no significant differences in overall accuracy rates ($p > 0.05$) between both sets of participants. Therefore, scores for these 38 participants were pro-rated for the following statistical analyses. Other missing values were managed by mean substitution. For all variables of interest, out-of-range values, plausible means, and standard deviations were examined. All variables of interest showed normal distributions with skewness values below three and kurtosis values below 10 (Kline, 2005).

Regression Assumptions. Prior to conducting each regression, diagnostic tests were completed to determine whether regressions were a valid procedure. Assumptions of singularity, linearity, homoscedasticity, absence of multicollinearity, and lack of unusual points (i.e., outliers, high leverage point, highly influential points) were verified. Singularity (i.e., independence of observations) was assumed as none of the independent variables were a combination of other independent variables (Tabachnick & Fidell, 2013). There was also independence of residuals, as indicated by a Durbin-Watson statistic that ranged from 1.92 – 2.20 across regression analyses.

Assumptions of linearity (i.e., predictor variables are collectively and independently linearly related to all outcome variables), homoscedasticity (i.e., equal residuals for all values of the outcome variables), and normally distributed errors (i.e., normal distribution of the residuals) were examined through residual scatter plots, with no violations found. Additionally, the assumption of collinearity was met as no independent variables were highly correlated and the collinearity statistics were all within acceptable ranges (i.e., tolerance values were all greater than 0.2 and Variance Inflation Factors were all within 1 and 10) (Tabachnick and Fidell, 2013; Hair et al., 1998). Finally, the absence of unusual points was verified (i.e., scores were less than three standard deviations from the mean, leverage values were below 0.20, and Cook's Distance values were above 1).

Accuracy, Confidence, and Indices of Monitoring Accuracy

The first goal of the current study was to measure and compare monitoring accuracy indices (i.e., calibration and resolution) *within* a facial emotion recognition domain. For comparison, monitoring accuracy indices were also measured and compared within a general knowledge domain.

Success Rates and Confidence. Mean success rates and confidence ratings are presented in Tables 5 and 6 for the general knowledge task and the emotional face recognition task respectively. Additionally, Figure 4 depicts mean success rates and confidence ratings across the general knowledge and emotional face recognition tasks. As designed to be, mean success rates in both tasks ($M_{generalknowledgesuccess}$ = 0.53; SD = 0.13; $M_{emotionalfacerecognitionsuccess}$ = 0.65, SD = 0.11) (allowed confidence variability above and below actual success rates. The small differences in accuracy allowed for comparison across tasks The two monitoring accuracy tests displayed adequate variability in success rates (general knowledge: 21% to 94%; emotion recognition: 35%

to 91%) and in confidence ratings (general knowledge: 27.9% to 96.4%; eyes: 43.7% to 100%). This variability allows for calculating resolution (Fleming & Lau, 2014).

Table 5

Descriptive Data for General Knowledge Task Dependent Variables (N = 136)

Dependent Measure	Mean (SD)	Potential Range	Observed Range	Skewness	Kurtosis
Pre-Task Ratings					
Pre-Task Confidence	5.80 (2.09)	0 – 10	1 – 10	- 0.01	- 0.73
Task Performance and Monitoring					
Success	0.53 (0.13)	0 – 1	.21 – .94	0.03	0.20
Confidence	0.72 (0.12)	0 – 1	.28 – .96	- 0.76	0.91
Bias Index[2]	0.19 (0.14)***	-1 – 1	-.15 – .48	- 0.20	- 0.12
Overconfidence Index[3]	0.20 (0.12)***	0 – 1	0 – .48	- 0.38	- 0.48
Resolution (Gamma)[4]	0.35 (0.28)***	0 – 1	0 – 1	0.46	- 0.79
Post-Task Ratings					
Post-Task Confidence	5.87 (1.88)	0 – 10	0 – 10	- 0.33	0.31
Task Difficulty	3.90 (2.10)	0 – 10	0 – 10	0.12	- 0.27
Effort Required	4.62 (2.07)	0 – 10	0 – 9	- 0.23	- 0.38
Feeling of Effort	4.60 (1.74)	0 – 10	0 – 9	- 0.48	0.89

*Asterisks indicate the significance of one sample t-tests for differences of bias, overconfidence, and resolution from zero.

*$p < .05$; **$p < .01$; ***$p < .001$

[2] Scores range from – 1 to 1. Scores above zero indicated overconfidence and scores below zero indicated underconfidence

[3] Scores range from 0 to 1. Higher score indicates poorer calibration and larger discrepancy between confidence and accuracy

[4] Scores range from 0 to 1. Higher score indicates better resolution and larger discrimination between correct and incorrect answers

Table 6

Descriptive Data for Emotional Face Recognition Task Dependent Variables (N = 136)

Dependent Measure	Mean (SD)	Potential Range	Observed Range	Skewness	Kurtosis
Pre-Task Ratings					
Pre-Task Confidence	7.15 (1.74)	0 – 10	3 – 10	-0.14	0.69
Task Performance and Monitoring					
Success	0.65 (0.11)	0 – 1	.35 – .91	0.07	-0.43
Confidence	0.76 (0.12)	0 – 1	.44 – 1.00	-0.19	-0.47
Bias[5] Index	0.11 (0.15)***	-1 – 1	-.26 – .45	0.02	-0.28
Overconfidence Index[6]	0.15 (0.11)***	0 – 1	0 – .45	-0.76	-0.03
Resolution (Gamma)[7]	0.29 (0.23)***	0 – 1	0 – .83	0.27	-1.01
Post-Task Ratings					
Post-Task Confidence	6.21 (2.38)	0 – 10	1 – 10	-0.34	-0.64
Task Difficulty	4.91 (2.58)	0 – 10	0 – 10	-0.26	0.81
Effort Required	5.32 (2.09)	0 – 10	0 – 10	-0.22	0.32
Feeling of Effort	4.69 (1.83)	0 – 10	0 – 10	0.20	1.02

*Asterisks indicate the significance of one sample t-tests for differences of bias, overconfidence, and resolution from zero.
*$p < .05$; **$p < .01$; ***$p < .001$.

[5] Scores range from – 1 to 1. Higher score indicated greater overconfidence and lower score indicates greater underconfidence
[6] Scores range from 0 to 1. Higher score indicates poorer calibration and larger discrepancy between confidence and accuracy
[7] Scores range from 0 to 1. Higher score indicates better resolution and larger discrimination between correct and incorrect answers

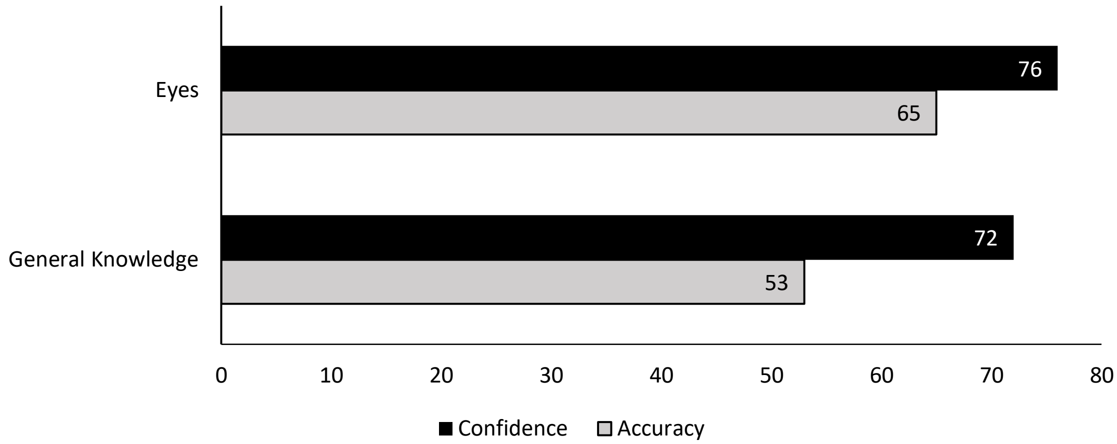


Figure 4. Mean accuracy and mean confidence across the emotion recognition and general knowledge tasks.

The relationship between accuracy and confidence was examined across monitoring accuracy tests. Pearson correlations between mean success rates and confidence ratings for the emotional face recognition task and general knowledge task are presented in Table 7. There was a moderate statistical relationship between success and confidence on the general knowledge task, r=.41, p<.001 and a small statistical relationship between success and confidence on the emotional face recognition task, r=.23, p<.01. These correlations indicated that increased success was associated with increased confidence on both tasks.

Table 7
Correlations among Success, Confidence, Calibration and Resolution for the Emotional Face Recognition Task and the General Knowledge Task (N= 136).

	Mean Success	Mean Confidence	Calibration	Resolution
Emotional Face Recognition				
Mean Success	--			
Mean Confidence	.23**	--		
Calibration a	.50***	-.46***	--	
Resolution b	.02	.03	-.04	--
General Knowledge				

Mean Success	--			
Mean Confidence	.41***	--		
Calibration [a]	.62***	-.38***	--	
Resolution [b]	.11	.24***	-.10	--

*p < .05; **p < .01; ***p < .001.

[a] Resistance to Overconfidence Index, where a higher score indicated better calibration.
[b] Resolution index, where a higher score indicated better discrimination between correct and incorrect response.

Calibration. To test the hypothesis that overconfidence exists in an emotional face recognition domain in addition to a general knowledge domain, calibration was examined for the general knowledge and emotion recognition tasks. Means and standard deviations for indices of calibration (i.e., Bias and Overconfidence) are presented in Tables 5 and 6 for the general knowledge task (M_{bias} = 0.19, SD = 0.14; $M_{overconfidence}$ = 0.20, SD = 0.12) and the emotion recognition task (M_{bias} = 0.11, SD = 0.15; $M_{overconfidence}$ = 0.15, SD = 0.11).

Bias Index. Figure 5 describes the proportion of participants who were overconfident, underconfident and perfectly calibrated on the general knowledge and emotion recognition tasks. For the general knowledge task, out of 136 participants, 121 participants were overconfident (i.e., had scores above zero), one participant was perfectly calibrated (i.e., had a score of zero) and 14 participants were underconfident (i.e., had scores below zero). For the emotion recognition task, out of 136 participants, 102 participants were overconfident (i.e., had scores above zero), four participants were perfectly calibrated (i.e., had scores of zero) and 30 participants were underconfident.

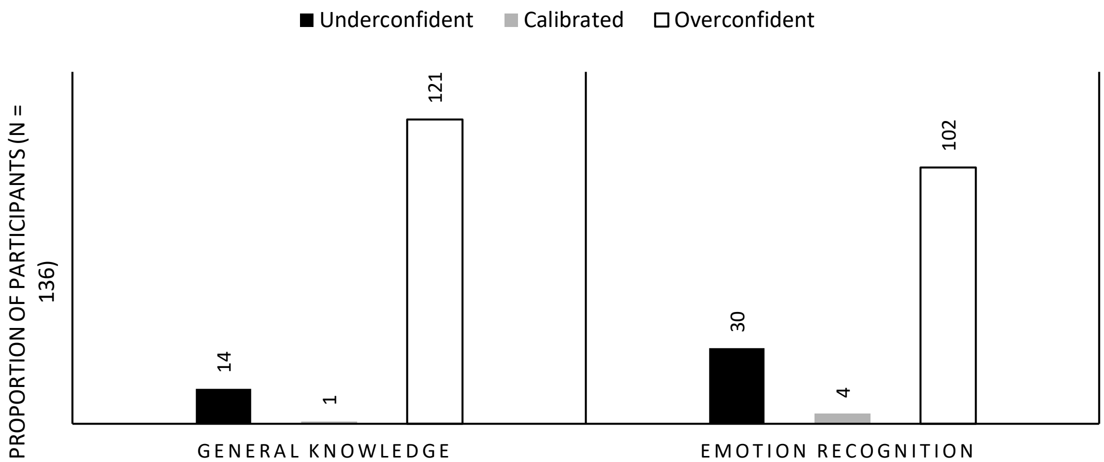


Figure 5. Proportion of participants (N = 136) who were underconfident, overconfident, and perfectly calibrated on the general knowledge and emotion recognition tasks.

Consistent with the study hypotheses, a significant overconfidence bias was found in both the general knowledge and emotional face recognition tasks. The significance of one-sample t-tests for differences of Bias from zero are marked by asterisks near the means in Tables 5 and 6. The Bias Index was significantly greater than zero for the general knowledge task $t(135) = 16.22$ $p<.001$ and emotional face recognition task $t(135) = 8.75$, $p<.001$, indicating significant overconfidence. Overconfidence bias is reflected in a Bias score above zero and can reach a maximum score of one, which would reflect 100% confidence and no accuracy. Mean Bias scores (for the general knowledge task, $M_{bias} = 0.19$, SD = 0.14 and the emotion recognition task, $M_{bias} = 0.11$, SD = 0.15) were at the lower end of the overconfidence range, suggesting that while there was overconfidence, the degree of overconfidence is relatively small. However, as illustrated in Figure 4, these findings suggested that participants on average were more confident than accurate in both general knowledge and emotion recognition domains, consistent with previous research. There were no significant sex differences in the Bias Index for general knowledge, $t(134) = 0.63$, $p=0.49$ or emotional face recognition, $t(134) = 0.09$, $p= 0.93$.

Resistance to Overconfidence Index. As previously described, the Bias Index cannot be examined in correlational analyses for investigations including variables where scores are unidirectional and range from zero to one. Therefore, the Resistance to Overconfidence Index, which is the inverse of the Overconfidence index, was used for all subsequent calibration analyses. The significance of one-sample t-tests for differences of the Overconfidence index scores from zero are marked by asterisks near the means in Tables 5 and 6 for the general knowledge task and emotion recognition task respectively. Mean Overconfidence index scores were significantly greater than zero for the general knowledge task, $t(135) = 79.93$, $p<0.001$ and the emotion recognition task, $t(135) = 91.91$, $p<0.001$. These findings indicated that overall confidence did not match accuracy on either test (although the miscalibration was towards the lower end of the range and relatively small). Additionally, there were no significant sex differences in this index for general knowledge, $t(134) = 0.72$, $p=0.47$ or emotional face recognition, $t(134) = -0.15$, $p=0.37$.

Resolution. To test the hypothesis that discrimination between correct and incorrect answers would occur for both general knowledge and emotional face recognition, resolution, measured by the Goodman-Kruskal Gamma correlation, was examined for the general knowledge and emotional face recognition tasks. Mean resolution is presented in Table 5 for the general knowledge task ($M_{resolution} = 0.35, SD = 0.28$) and Table 6 for the emotion recognition task ($M_{resolution} = 0.29, SD = 0.23$). Figure 6 depicts mean confidence for incorrect and correct responses across the general knowledge task and the emotion recognition task. On the emotion recognition task, mean confidence was 78.18% (*SD =17.02*) for correct responses and 72.36% (*SD= 18.35*) for incorrect responses. On the general knowledge task, mean confidence was 78.32% (*SD= 26.16*) for correct responses and 65.27% (*SD = 26.00*) for incorrect responses.

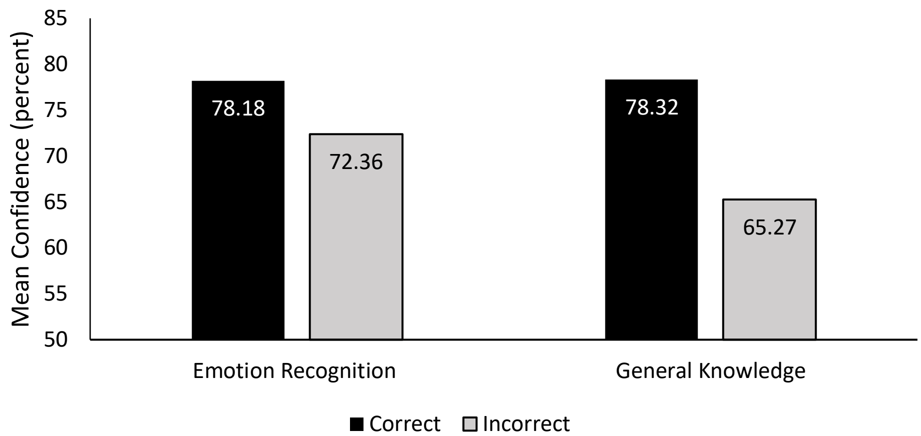


Figure 6. Mean confidence (percent) for correct and incorrect choices on the emotion recognition and general knowledge tasks.

Significant resolution effects were found for both tasks, indicating that participants discriminated successfully between correct and incorrect responses. The significance of one-sample t-tests for differences of resolution from zero are marked by asterisks near the means in Tables 5 and 6. Mean resolution was significantly greater than zero on the general knowledge task $t(135) = 14.85$, $p < .0001$ and the emotion recognition task, $t(133) = 14.39$, $p < 0.001$. There were no significant sex differences in resolution for general knowledge, $t(134) = 1.58$, $p = .12$ or emotion recognition, $t(132) = -0.75$, $p = .46$. Thus, while participants were generally overconfident on general knowledge and emotion recognition tasks as indicated by the Bias Index, they tended to be less confident for incorrect choices and more confident for correct choices.

Calibration versus Resolution. To test the hypothesis that calibration and resolution are separable indices of monitoring accuracy that would not be correlated within domains, associations between calibration, measured by the Resistance to Overconfidence Index, and resolution, measured by the Goodman-Kruskal Gamma correlation, were examined *within* each experimental task (see Table 7 above); as expected, no significant relationship was found between calibration and resolution within domains, *p*s>.05. These findings suggested that calibration was not associated with successful discrimination between incorrect and correct responses for either general knowledge or emotion recognition domain.

Of note, across both the general knowledge and emotional face recognition tasks, mean success rates were positively correlated with calibration but not resolution (see Table 7 above). Specifically, there was a large statistical relationship between success and calibration for the emotional face recognition task $r(136)=.50$, $p<.001$, a large statistical relationship between success and calibration for the general knowledge task, $r(136)=.62$, $p<.001$, and no statistical relationships between success and resolution for either task, $p>.05$. These results indicated that while better performance was correlated with greater Resistance to Overconfidence Index scores, performance was unrelated to the ability to discriminate between correct and incorrect choices. This finding further supports the separability of calibration and resolution.

Domain-Generality versus Domain-Specificity: Monitoring Accuracy Across Domains

A second goal of the current study was to explore the question of whether indices of monitoring accuracy remain stable *across* domains or rather fluctuate as a function of domain knowledge. Therefore, calibration and resolution were examined *across* general knowledge and emotional face recognition domains.

To test the hypothesis that calibration would be domain-general, Pearson correlations were conducted to measure associations between Resistance to Overconfidence scores on the general knowledge task and Resistance to Overconfidence scores on the emotional face recognition task. A small statistical relationship was found between calibration on the general knowledge task and calibration on the emotional face recognition task, $r(136)= 0.23$, $p<0.01$.

This correlation indicated that better calibration on one task was associated with better calibration on the other task.

To test the hypothesis that resolution would be domain-specific, Pearson correlations were used to measure the relationship between resolution on the general knowledge task and resolution on the emotion recognition task. Resolution for general knowledge was not significantly correlated with resolution for emotion recognition $r(136) = -.07$, $p > .05$. In contrast to calibration, this result suggested that reliable discrimination between correct and incorrect responses on one task was not associated with reliable discrimination between correct and incorrect responses on the other task.

Success, Confidence, and Monitoring Accuracy Correlates

A third goal of the current study was to explore potential correlates of accuracy, confidence, and monitoring accuracy, particularly in the rarely examined domain of emotional face recognition, including: (1) cognitive abilities (i.e., working memory and intellectual abilities), (2) task-specific judgments (i.e., pre- and postdictive confidence, aggregated ratings of task difficulty, effort required, and feeling of effort) (3) self-perceptions (i.e., perceived empathic ability; self-reported social anxiety) and (4) cognitive emotion regulation skills.

Cognitive Abilities. Table 8 contains raw means (and standard deviations) for the working memory task scores, Shipley vocabulary test scores, and Shipley block pattern test scores. The scores were normally distributed. Overall mean scores suggest that there were no floor or ceiling effects on these tests.

Table 8
Descriptive Statistics for Cognitive Ability Measures

	Mean (SD)	Potential Range	Observed Range	Skewness	Kurtosis

Working Memory Total Raw Score	32.10 (4.09)	0 – 42	20 – 42	-0.20	0.07
Shipley- 2 Vocabulary Raw Score	25.92 (4.35)	0 – 40	16 – 37	-0.1	-0.10
Shipley-2 Block Patterns Raw Score	16.18 (4.92)	0 – 26	5 – 26	0.01	-0.79
Intelligence Z-score Composite	42.10 (6.94)	0 – 66	21 – 59	-0.01	-0.32

Note. N = 136

To test the hypothesis that cognitive abilities would be positively correlated with monitoring accuracy indices across domains of facial emotion recognition and general knowledge, relationships between cognitive abilities, success, confidence, and monitoring accuracy indices (i.e., calibration and resolution) were examined for both tasks. Table 9 displays Pearson correlations between the cognitive ability scores, success, confidence and monitoring accuracy indices on the general knowledge and emotional face recognition tasks.

Table 9

Correlations Between Cognitive Abilities and Success, Confidence, Calibration, and Resolution

	Shipley- 2 Vocabulary Raw Score	Shipley-2 Block Patterns Raw Score	Intelligence Z-score Composite	Working Memory Total Raw Score
Emotional Face Recognition				
Mean Success	.30***	.12	.28***	.25***
Mean Confidence	.05	-.09	-.02	-.06
Calibration a	.19*	.03	.15	.23**
Resolution b	-.12	-.18*	-.21*	-.00
General Knowledge				
Mean Success	.41***	.29**	.48***	.38***
Mean Confidence	.37***	.16	.36***	.11
Calibration a	.10	.12	.14	.22**
Resolution b	.17	.22**	.26**	.13

p < .05; ***p* < .01; ****p* < .001
[a] Resistance to Overconfidence Index, where a higher score indicated better calibration.
[b] Resolution index, where a higher score indicated better discrimination between correct and incorrect response.

Cognitive Abilities and Monitoring Accuracy. As expected, working memory scores were significantly and positively correlated with Resistance to Overconfidence on both tasks. Specifically, there was a small positive relationship between working memory performance and calibration for emotional face recognition, $r(136)=.23$, $p<.01$ and a small positive relationship between working memory performance and calibration for general knowledge, $r(136)=.22$, $p<.01$. These findings suggested that higher working memory performance was related to lower overconfidence in both domains, in line with previous research (Bruine de Bruin et al., 2007; Stanovich & West, 1998).

However, contrary to expectations, working memory scores were not significantly correlated with resolution for either task, $p>.05$. Therefore, higher performance on the working memory task was consistently unrelated to reliable discrimination between incorrect and correct responses. These findings suggested that working memory was not related to resolution in an emotional face recognition domain or a general knowledge domain.

Although intellectual ability composite scores correlated with success on both tasks, no significant relationships were found between the intellectual ability composite scores and calibration, contrary to expectations. There was a small positive relationship between verbal ability scores and calibration on the emotion recognition task, $r(136)=.19$, $p<.05$. Verbal ability was not significantly correlated with confidence for emotional face recognition task, but it was moderately correlated with success $r(136)=.30$, $p<.001$ suggesting that the relationship between verbal ability and calibration on the emotional face recognition task may have varied according to proficiency in emotional face recognition.

Relationships between intellectual ability and resolution varied across tasks. The intellectual ability composite score was moderately and positively correlated with resolution on the general knowledge task, $r(136)=.26$, $p<.01$. However, contrary to expectations, the intellectual ability composite score was negatively correlated with resolution on the emotional face recognition task. Specifically, a small negative relationships was found between intellectual ability composite scores and resolution, $r(136)=-.21$, $p<.05$. These correlations indicated that higher intellectual ability was associated with greater discrimination between correct and incorrect responses in general knowledge, as expected, but lower discrimination between correct and incorrect responses in emotional face recognition.

Aggregated Task-Specific Judgments. The means (and standard deviations) for predictive judgments of confidence and postdictive judgments of confidence, difficulty, and effort are presented in Table 5 for the general knowledge task and Table 6 for the emotional face recognition task. As previously described, pre- and postdictive confidence judgments represented aggregated judgments regarding the ability to answer task-related questions (i.e., “How confident are you in your ability to correctly solve general knowledge questions?”), rather than item-by-item confidence. The means on the aggregated confidence ratings ranged in the 4-6 range on the scales, indicating that none of the tasks were rated as extremely difficult, extremely unpleasant, or requiring extreme effort.

To test the hypothesis that greater calibration and resolution would be correlated with higher predictive and postdictive confidence and lower task difficulty and effort ratings, Pearson correlations among aggregated task-specific judgments, success, confidence, and monitoring accuracy indices were examined. Pearson correlations are presented in Table 10.

Table 10.
Correlations among pre-task confidence, post-task ratings, calibration and resolution for the General Knowledge and Emotion Recognition Tasks

	Predictive Confidence	Postdictive Confidence	Task Difficulty	Effort Required	Feeling of Effort
General Knowledge					
Success	.07	.25**	-.15	-.12	-.04
Confidence	.31***	.57***	-.48***	-.25**	-.17*
Calibration [a]	-.24**	-.21*	.21*	.10	.13
Resolution [b]	.10	.19*	-.21*	-.16	-.07
Emotional Face Recognition					
Success	.11	.02	-.09	-.19*	-.13
Confidence	.35***	.55***	-.34***	-.22*	-.19*
Calibration [a]	-.21*	-.37***	.17*	-.05	-.07
Resolution [b]	.21**	.07	-.06	.05	-.10

*p<.05; **p<.01; ***p<.001
[a] Resistance to Overconfidence Index, where a higher score indicated better calibration.
[b] Resolution index, where a higher score indicated better discrimination between correct and incorrect response.

Predictive and Postdictive Confidence. In some contrast to initial hypotheses that predicted calibration would be positively correlated with ratings of predictive and postdictive confidence, we found negative relationships between calibration and ratings of predictive and postdictive confidence. Specifically, a small negative correlation was found between Resistance to Overconfidence scores and predictive confidence for general knowledge $r(136)=-.24$, p<.01 and Resistance to Overconfidence scores and predictive confidence for emotional face recognition $r(136)=-.21$, p<.05. These findings suggested that higher confidence before a test was related to more overconfidence in each domain. Moreover, Resistance to Overconfidence was negatively correlated with post-task confidence for both domains. Specifically, a small negative relationship was found between Resistance to Overconfidence and post-task confidence for general knowledge, *r(136)*=.21, p<.05 and a moderate negative relationship was found between Resistance to Overconfidence and post-task confidence for emotional face recognition *r(136)*=-.37, *p*<.001. These correlations indicated that greater overconfidence on a test was related to higher confidence after a test. Interestingly, postdictive confidence was moderately, positively correlated with success on the general knowledge task, *r(136)*=.25, p<.01, but no significant relationship was found between postdictive confidence and success for the emotional face recognition task, *r(136)*=.02, p>.05. This suggested that postdictive confidence was related to increased proficiency on the general knowledge task but it was not related to greater proficiency on the emotion recognition task.

The relationship between resolution and ratings of predictive and postdictive confidence varied depending on the domain. Predictive confidence was positively correlated with resolution *r(136)*=.21, *p* <.01, as expected, but only for the facial emotion recognition task. Therefore, the higher the predictive confidence for facial emotion recognition, the better the discrimination between correct and incorrect answers on the facial emotion recognition task. There was a small significant relationship between resolution and postdictive confidence, but only for general knowledge, *r(136)*=.19, *p*<.05. Therefore, the higher the discrimination between correct and incorrect answers on the general knowledge task, the higher the confidence in performance after completing the task. Overall, resolution for emotional face recognition positively correlated with predictive confidence while resolution for general knowledge positively correlated with postdictive confidence.

Task Difficulty, Effort, Feeling of Effort. There was a small positive relationship between calibration and ratings of task difficulty for the general knowledge task, *r(136)*=.21, *p*<.05 and the emotional face recognition task *r(136)*=.17, *p*<.05. The correlations indicated that ratings of task difficulty increased as calibration scores increased, which was in contrast to initial hypotheses that predicted a negative relationship would occur between task difficulty and calibration. Contrary to hypotheses, required effort and feeling of effort ratings (where a higher score indicates less pleasant) were not significantly correlated with calibration, *p*s>.05. These results indicated that with better calibration, tasks were judged to be more difficult, but judgments of effort did not vary accordingly.

In some contrast to calibration, resolution was significantly associated with task difficulty, but only on the general knowledge task. There was a small negative correlation between task difficulty and resolution for the general knowledge task *r*=-.21, *p*<.05. This correlation suggested that the general knowledge task was judged to be less difficult as discrimination between correct and incorrect responses increased. However, while judgments of task difficulty increased with decreased resolution, judgments of effort did not vary accordingly, *r*=.10, *p*>.05.

Overall, better calibration correlated with lower predictive and postdictive confidence ratings and higher task difficulty ratings within both facial emotion recognition and general knowledge domains. In contrast, better resolution correlated with higher predictive and postdictive ratings and these relationships varied within domains; resolution for facial emotion recognition correlated with higher predictive confidence whereas resolution for general knowledge correlated with higher postdictive confidence. Required effort and discomfort ratings did not correlate with resolution or calibration.

Self-Perceptions and Cognitive Emotion Regulation. The means (and standard deviations) and normality estimates for self-perception variables including perceived empathic ability (measured by the Toronto Empathy Questionnaire, social anxiety (measured by the Social Anxiety in University Students Scale), and differentiation of emotions (measured by the Differentiation of Emotions Scale) as well as cognitive emotion regulation (measured by the Cognitive Emotion Regulation Questionnaire-Short) are displayed in Table 11. Responses were normally distributed, with no floor or ceiling effects. Relationships between responses on these scales and performance on the tests of cognitive ability (i.e., Shipley-2 vocabulary and block patterns; working memory) were examined using Pearson correlations. There were no significant correlations found, *p*s>.05. Relationships among the scales were examined using Pearson correlations. Social anxiety scores decreased with increased perceived empathic ability r=-.27, p <.05, differentiation of emotions r =-.37, p<.001, and use of adaptive cognitive coping skills r=-.36, p <.001. These correlations are consistent with previous research (e.g., Kaur and Kaur, 2015; Ertekin, Koyuncu, Aslantas, & Özyıldırım, 2015; Martin & Dahlen, 2005; Knight, 2019; Morrison, Mateen, Brozovich, Zaki, Golden et al., 2016).

Table 11
Descriptive Statistics for Cognitive Emotion Regulation, Self-Perceptions, and Social-Emotional Challenges (N = 136)

Scale	M (SD)	Potential Range	Observed Range	Skewness	Kurtosis
Self-Perception Measures:					
Cognitive Emotion Regulation[a] composite	47.15 (7.35)	12 – 72	20 – 64	-0.50	1.05
Empathy	74.21 (10.31)	16 – 96	46 – 93	-0.30	-0.41
Differentiation of emotions	56.10 (9.46)	14 – 84	31 – 79	0.12	-0.25
Social anxiety	41.86 (10.09)	15 – 90	22 – 76	-0.68	0.75
Behavioural Correlate:					
Social-Emotional challenges	1.56 (0.71)	1 – 4	1 – 4	1.0	0.67

[a]the Cognitive Emotion Regulation score is a composite sum of the self-blame (inverse), other-blame (inverse), catastrophizing (inverse), positive reappraisal, refocus on planning and positive refocusing subscale scores.

Correlations between Self-perceptions, Confidence, and Monitoring Accuracy. To explore self-perceptions as correlates of monitoring accuracy in a social domain**,** relationships between self-perceptions, confidence, and monitoring accuracy indices were examined for the emotional face recognition task. For comparison, these relationships were also examined for the general knowledge task although they were not expected to be significant in this domain. Pearson correlations are presented in Table 12.

Table 12
Correlations between emotional face recognition task dependent variables and measures of cognitive emotion regulation, self-perceptions, and social-emotional challenges (N = 136)

	Cognitive Emotion Regulation	Empathy	Differentiation of Emotions	Social Anxiety	Social-Emotional Challenges
Emotional Face Recognition					
Mean Success	.03	.23**	.13	-.13	-.08
Mean Confidence	.18*	.11	.23**	-.39***	-.25**
Resistance to Overconfidence[a]	.08	-.04	.12	-.08	.05
Resolution[b]	-.07	.07	.05	-.21*	-.08
General Knowledge					
Mean Success	.13	.02	.01	-.04	-.02
Mean Confidence	.15	.07	.17	-.23**	-.26**
Resistance to Overconfidence	.01	-.02	.12	-.16	-.09
Resolution	-.11	-.09	-.17	.08	-.06

*$p < .05$; **$p < .01$; ***$p < .001$
[a] Resistance to Overconfidence Index, where a higher score indicated better calibration.
[b] Resolution index, where a higher score indicated better discrimination between correct and incorrect response.

The only self-perception variable to correlate with confidence or monitoring accuracy was self-reported social anxiety. A moderate negative correlation was found between self-reported social anxiety and confidence for emotional face recognition, r=-.39, p<.001. A small negative correlation was found between self-reported social anxiety and resolution for emotional face recognition r=-.21, p<.05. These findings indicated that the greater the self-reported social anxiety, the lower the confidence and the discrimination between correct and incorrect responses on the emotional face recognition task. Self-reported social anxiety was not correlated with success on the emotional face recognition task, suggesting that the relationship between resolution and perceived social anxiety is not attributed to proficiency in emotional face recognition. Confidence for general knowledge was also negatively correlated with self-reported social anxiety r(136)=-.23, p<.01, but no relationship was found between self-reported social anxiety and monitoring accuracy on the general knowledge task, r(136)=.02, p>.05. Thus, while self-reported social anxiety correlated with confidence regardless of domain-content, the relationship between self-reported social anxiety and resolution was unique to emotional face recognition. Contrary to hypotheses, no significant relationship was found between confidence for emotional face recognition and perceived empathic ability, r(136) =.11, p>.05. In addition, no significant relationships were found between resolution for emotional face recognition and perceived empathic ability or differentiation of emotions, *ps*>.05. Finally, no significant relationships were found between calibration for emotional face recognition and self-perceptions, p>.05, indicating that none of the self-perception variables examined in the current study was associated with calibration.

Cognitive Emotion Regulation. To test the hypothesis that monitoring accuracy in a social domain would be positively correlated with adaptive use of cognitive emotion regulation strategies, relationships between responses on the Cognitive Emotion Regulation Questionnaire-Short and calibration and resolution for emotional face recognition were examined (see Table 12 above). There were no significant relationships between cognitive emotion regulation composite scores and calibration or resolution for the emotional face recognition task (or the general knowledge task), *p*s$>.05$. Thus, contrary to expectations, these findings indicated that more adaptive use of cognitive emotion regulation skills was not related to better calibration or discrimination between correct and incorrect choices for emotional face recognition.

Predictors of Monitoring Accuracy

Correlates of calibration and resolution were further explored by using simultaneous regression analyses to identify unique predictors of monitoring accuracy. Due to the pattern of correlations found above, the following variables were explored as predictors of monitoring accuracy on the emotional face recognition task: (1) working memory performance, (2) intelligence test scores, (3) task-specific judgments (predictive confidence, postdictive confidence, and post-task difficulty ratings) and (4) self-perceptions of social anxiety. For comparison, working memory performance, intelligence test scores, and task-specific judgments were also explored as correlates of monitoring accuracy on the general knowledge task (perceived social anxiety was not hypothesized to predict general knowledge monitoring accuracy nor did it correlate with general knowledge monitoring accuracy and, thus, was not included).

Predicting Calibration. To identify unique predictors of calibration, simultaneous regressions were conducted with calibration, measured by the Resistance to Overconfidence Index, as the dependent variable. Simultaneous regression analyses to predict calibration are summarized in Table 13.

Table 13.
Simultaneous Regression Results for Resistance to Overconfidence

	B	*SE*B	β	t	Variance explained
Criterion Variable: Resistance to Overconfidence Index[a] on the Emotional Face Recognition Task					
Working memory raw score	.005	.002	.19	2.30*	4%
Intelligence raw composite z-score	.005	.006	.08	0.84	<1%
Predictive Confidence	-.002	.006	-.06	-0.60	<1%
Postdictive Confidence	-.018	.005	-.34	-3.40**	8%
Task Difficulty	.007	.004	.17	1.70	2%
Social Anxiety	-.000	.001	-.03	-0.38	<1%
Overall Regression: F (6, 129) = 4.97***					
Multiple R = 0.43					
Multiple R^2 = 0.19					
Adjusted R^2= .15					
Criterion Variable: Resistance to Overconfidence Index[a] on the General Knowledge Task					
Working memory raw score	.007	.002	.23	2.74**	5%
Intelligence raw composite z-score	.012	.007	.15	1.82	3%
Pre-Task Confidence	-.006	.006	-.11	-1.12	<1%
Post-Task Confidence	-.006	.007	-.09	-.84	<1%
Task Difficulty	.011	.005	.21	2.23*	4%
Overall Regression: F (5, 130) = 5.02***					
Multiple R = 0.40					
Multiple R^2 = 0.16					
Adjusted R^2 = 0.13					

Note. *B* = unstandardized regression coefficient; *SE*B = standard error of the coefficient; β = standardized coefficient.
[a] A higher score indicated better calibration.
*p < .05; **p < .01; ***p < .001

Predicting Calibration for Emotional face recognition. A simultaneous multiple regression was conducted to predict the Resistance to Overconfidence Index for emotional face recognition from cognitive abilities, task-specific judgments (i.e., predictive confidence, postdictive confidence, and task difficulty) and self-perceptions of social anxiety. The model was significant R^2=.19, F(6, 129)=4.97*** p<.001, adjusted R^2=.15. Working memory emerged as a unique, significant predictor of Resistance to Overconfidence β=.19, p< .05, explaining 4% of the unique variance, as did postdictive confidence β =-.34, p<.01, explaining 8% of the unique variance.

Predicting Calibration for General Knowledge. For comparison, a simultaneous multiple regression was also conducted to predict the Resistance to Overconfidence Index on the general knowledge task from cognitive abilities and task-specific judgments (i.e., predictive confidence, postdictive confidence, and task difficulty). The model was significant R^2=.17, F(6,128)=3.68, p<.001, adjusted $R^2=0.13$. The results indicated that working memory performance significantly predicted the Resistance to Overconfidence Index β=.24, p<.01 and uniquely explained 5% of the variance. Task difficulty also emerged as a significant predictor, β=.21, p<.01, explaining 4% of the unique variance. No other variables emerged as significant predictors of the Resistance to Overconfidence Index on the general knowledge task.

Overall, working memory performance emerged as a significant predictor of Resistance to Overconfidence scores across domains. Postdictive confidence emerged as an additional predictor of Resistance to Overconfidence scores in an emotional face recognition domain while task difficulty emerged as an additional predictor of Resistance to Overconfidence in a general knowledge domain.

Predicting Resolution. To identify unique predictors of resolution, simultaneous regressions were conducted with the resolution index, measured by the Goodman-Kruskal Gamma correlation, as the dependent variable. Simultaneous regression analyses to predict resolution are summarized in Table 14.

Table 14.
Simultaneous Regression Results for Resolution

	B	*SE*$_B$	*β*	t	Unique Variance explained
Criterion Variable: Resolution[a] on the Emotional face recognition Task					
Working memory raw score	.004	.005	.09	0.99	<1%
Intelligence composite z-score	-.041	.014	-.26	-3.01**	7%
Predictive Confidence	.031	.013	.23	2.43*	5%
Postdictive Confidence	-.015	.011	-.16	-1.49	2%
Task Difficulty	-.003	.008	-.03	-0.32	<1%
Social Anxiety	-.005	.002	-.21	-2.31*	4%
Overall Regression: F (6, 127) = 3.36**					
Multiple R = .37					
Multiple R^2 = .14					
Adjusted R^2 =.10					
Criterion Variable: Resolution[a] on the General Knowledge Task					
Working memory raw score	.004	.006	.055	0.64	<1%
Intelligence composite z-score	.041	.017	.212	2.48*	5%
Pre-Task Confidence	.000	.014	-.002	-.016	<1%
Post-Task Confidence	.018	.017	.123	1.10	1%
Task Difficulty	-.013	.013	-.097	-1.01	<1%
Overall Regression: F (5, 130) = 3.03*					
Multiple R = .32					
Multiple R^2 = .10					
Adjusted R^2 = .07					

*$p < .05$; **$p < .01$; ***$p < .001$
[a] A higher score reflects better discrimination

Predicting Resolution for Emotional Face Recognition. A simultaneous regression was conducted to predict resolution from working memory total scores, Shipley-2 intelligence scores, task-specific judgments (predictive confidence, postdictive confidence, and task-difficulty), and self-perceptions of social anxiety. The model was significant R^2=.14, F(6,127)=3.36, p<.01, adjusted R^2=.10. Intelligence, predictive confidence, and social anxiety emerged as unique predictors of resolution. Intelligence negatively predicted resolution, β=.26, p<.05, accounting for 7% of the variance. Pre-task ratings positively predicted resolution, β=.23, p<.05, explaining 5% of the variance. Self-reported social anxiety negatively predicted resolution, β=-.21, p<.0, explaining 4% of the variance in resolution.

Predicting Resolution for General Knowledge. For comparison, a simultaneous regression was conducted to predict resolution on the general knowledge task from working memory total scores, intelligence composite z-scores, and task specific judgments (predictive confidence, postdictive confidence, and task difficulty). The model for general knowledge was significant R^2=.11, F(5,130)=2.31, p<.05. Intelligence composite scores (i.e., verbal and non-verbal ability) significantly predicted resolution on the general knowledge task β = .21, p<.05, explaining 5% of the unique variance. No other variables emerged as significant predictors of resolution on the general knowledge task.

Overall, the results of the regression analyses indicated that the most consistent predictors of both calibration and resolution were cognitive abilities. Specifically, either the Shipley intelligence scores, or working memory total score were significant predictors of calibration or resolution. Working memory performance consistently positively predicted Resistance to Overconfidence on the general knowledge and emotion recognition tasks. However, while intelligence scores positively predicted resolution in general knowledge, intelligence negatively predicted resolution for emotional face recognition. Postdictive confidence and predictive confidence emerged as significant predictors of monitoring accuracy, but only for emotional face recognition. Task difficulty also emerged as a significant predictor, but only for general knowledge. In terms of self-perceptions, social anxiety emerged as a significant predictor of resolution, but not calibration, for emotional face recognition.

Behavioural Correlate

The final goal was to explore monitoring accuracy on the emotional face recognition task in relation to a behavioural correlate. Relationships between self-reported social, emotional, or behavioural challenges and monitoring accuracy for emotional face recognition were examined using Pearson correlations (see Table 13 above). Contrary to hypotheses, lower self-reported social-emotional challenges was not significantly correlated with calibration or resolution on the emotional face recognition task, *p*s>.05. Of note, there was a moderate negative correlation between confidence and self-reported social-emotional challenges for the emotional face recognition task, $r(136)=-.25$, $p<.01$ and no significant relationship between self-reported social-emotional challenges and accuracy on the facial emotion recognition task, $r(136)=-.08$, $p>.05$. suggesting that self-reported social-emotional challenges was correlated with confidence for emotional face recognition but not accuracy.

Predicting Self-Reported Social-Emotional Challenges. To explore whether monitoring accuracy indices would predict a behavioural correlate, a simultaneous regression analysis was conducted to identify unique predictors of self-reported social-emotional challenges. Cognitive abilities, indices of monitoring accuracy on the emotional face recognition task, the self-perceptions scores and cognitive emotion regulation scores were entered as predictors of self-reported social-emotional challenges. The results of this regression analysis appears in Table 15.

The regression model did significantly predict social-emotional challenges, $F(5, 121)=5.35$, $p<.001$. However, the cognitive emotion regulation composite score was the sole significant predictor of self-reported social-emotional challenges, $\beta= 0.42$, $p< 0.001$, accounting for 13% of the variability. Contrary to expectations, resolution and calibration did not emerge as significant predictors of self-reported social-emotional challenges in university students.

Table 15

Simultaneous Regression Results for Self-Reported Social-Emotional Challenges

	B	SE_B	β	t	Variance explained
Working memory raw score	.01	.02	.06	.77	<1%
Intelligence raw composite z-score	.00	.04	.00	.02	<1%
Cognitive emotion regulation composite	-.04	-.01	-.42	-4.54***	13%
Perceived empathic ability	-.00	-.01	-.04	-.52	<1%
Differentiation of emotions	-.01	-.01	-.07	-.77	<1%
Social Anxiety	-.17	-.24	-.06	-.67	<1%
Resolution	-.04	-.54	-.01	-.07	<1%
Calibration					
Overall Regression: F (7,121) = 5.35***					
Multiple R = 0.48					
Multiple R^2 = 0.24					
Adjusted R^2 = 0.19					

**p* < .05; ***p* < .01; ****p* < .001

[a] Resistance to Overconfidence Index, where a higher score indicated better calibration.

[b] Resolution index, where a higher score indicated better discrimination between correct and

Discussion

The current study extended metacognitive and cognitive science research by examining monitoring accuracy in the domain of emotional face recognition. Participants viewed images of facial expressions, identified the emotion in each image, and rated their confidence in each answer provided. Monitoring accuracy was examined using two different indices: calibration and resolution. As expected, confidence was not well matched with performance, with a tendency towards overconfidence rather than underconfidence (index of calibration). Consistent with Kelly and Metcalfe's (2011) findings, this study also showed that discrimination between correct and incorrect answers (index of resolution) occurs in emotional face recognition, with individuals generally being more confident when they correctly identify emotions based on images of facial expressions than when they incorrectly identify emotions based on images of facial expressions. In addition, a general knowledge domain was included to better understand metacognitive processes across domains. Calibration in emotional face recognition was correlated with calibration in general knowledge, offering further evidence that calibration is domain-general. In contrast, resolution in emotional face recognition was not correlated with resolution in general knowledge, suggesting that resolution may be domain-specific. In addition, resolution and calibration were not correlated with one another within tasks, suggesting that resolution and calibration are separate constructs, a novel contribution of the current study. Several predictors of monitoring accuracy were explored, including cognitive abilities, task-specific judgments, and self-perceptions. Discrimination between correct and incorrect emotions based on images of facial expressions was negatively correlated with responses on the Social Anxiety in University Students Questionnaire, indicating that better discrimination related to lower perceptions of social anxiety. Main findings are presented in Table 16.

Table 16.
Main Aims and Findings

Aims		Hypotheses	Results: General Knowledge	Results: Emotion Recognition
Indices of Monitoring Accuracy	*Hypothesis 1*	Overconfidence in both tasks	√	√
		Resolution greater than zero in both tasks	√	√
		Calibration and resolution scores not correlated within tasks	√	√
Domain-Generality vs Domain-Specificity	*Hypothesis 2*	Calibration scores correlated across tasks	√	√
	Hypothesis 3	Resolution scores not correlated across tasks	√	√
Success, Confidence, and Monitoring Accuracy Correlates	*Hypothesis 4*	Cognitive abilities correlated with better calibration	√	√
		Cognitive abilities correlated with better resolution	√	Lower Resolution
	Hypothesis 5	Better calibration correlated with:		
		• Higher predictive confidence	Lower	Lower
		• Higher postdictive confidence	Lower	Lower
		• Lower Task difficulty	Higher	Higher
		• Lower Required effort		
		• Lower Feeling of effort		
		Better resolution correlated with:		
		• Higher predictive confidence		
		• Higher postdictive confidence		Higher
		• Lower difficulty	Higher	
		• Lower Required effort		
		• Lower Feeling of effort		

	Hypothesis 6	Better calibration and resolution correlated with: • Higher cognitive emotion regulation • Higher differentiation of emotion • Lower social anxiety • Higher perceived empathic ability	N/A	Resolution and Social Anxiety Only
	Hypothesis 7	Cognitive abilities and task-specific judgments predict calibration scores	Cognitive Abilities Only	Cognitive Abilities, Higher Postdictive Confidence
		Cognitive abilities and task-specific judgments predict resolution scores	Cognitive Abilities, Higher Task Difficulty	Cognitive Abilities, Higher Predictive Confidence
		Self-perceptions predict calibration and resolution for emotional face recognition	N/A	Social Anxiety Predicts Resolution Only
Behavioural correlates	*Hypothesis 8*	Self-reported social emotional challenges is predicted from emotional face recognition calibration and resolution.	N/A	

Note. N/A represents non-applicable analyses; a checkmark represents confirmed hypotheses.

Monitoring Accuracy in a Social Domain: Comparing Calibration and Resolution

The first aim of the current study was to measure and compare two indicators of accurate monitoring on an emotional face recognition task: resolution and calibration. To understand monitoring accuracy indices in different domains, resolution and calibration were also measured and compared on a general knowledge task. Mean accuracy was significantly correlated with mean confidence on both tasks. According to a within-subjects Goodman-Kruskal Gamma correlation for a measure of resolution, confidence was also generally higher for correct answers and lower for incorrect answers on both tasks. Resolution was not significantly correlated with accuracy, suggesting that proficiency on a task was not related to the ability to discriminate between correct and incorrect answers on a task. The current finding that resolution occurred on an emotional face recognition task is consistent with Kelly & Metcalfe's study (2011) and, as hypothesized, indicate individuals can generally discriminate between correct and incorrect inferences on the emotional face recognition task.

According to an index of calibration (i.e., Bias), confidence was not well matched with performance on either task. Bias was towards the lower end of the overconfidence range, with a tendency towards overconfidence rather than underconfidence. Therefore, participants' confidence in their performance was generally greater than their actual performance, which is consistent with previous research showing overconfidence in many domains of functioning, including general knowledge, mathematics, reading comprehension, and eye-witness memory (Denning, Heath, and Suls, 2004; Bruine de Bruine et al., 2007; Lichtenstein and Fischhoff, 1977; Stanovich et al., 2016; Glenberg & Epstein, 1987; Lin & Zabrucky, 1998; Ackerman & Zalmanov, 2012; Sporer et al.,1995; Wells, Olsen & Charman, 2002).

Calibration, measured by the Resistance to Overconfidence Index, was not significantly correlated with resolution for either task. Therefore, as expected, the ability to judge overall performance was not correlated with the ability to discriminate between correct and incorrect performance. This lack of correlation between calibration and resolution within tasks was also found by

Dentakos, Saoud, Ackerman, and Toplak (2019) for a financial task in an experimental paradigm similar to the one currently used. Current results reinforce previous claims that calibration and resolution are conceptually different and empirically separable indicators of monitoring accuracy (Ais et al., 2016; Koriat et al. 2002; Maki et al., 2005; Thiede et al., 2015).

Current findings support the conceptualization of resolution and calibration as complementary monitoring accuracy indices (see Ackerman, Parush, Nassar, & Shtub, 2016, for a review). Based on metacognitive theories, calibration (i.e., degree of fit between confidence and accuracy) is an index that measures the self-monitoring process that is expected to influence the overall amount of time that people invest into understanding a situation. Resolution (i.e., discrimination) is also a self-monitoring process, but this construct is expected to influence which parts of a situation people choose to allocate more time to understanding. Calibration and resolution are distinct indices that measure distinct monitoring accuracy abilities. This distinction between calibration and resolution is important for targeting metacognitive abilities in academic domains (see Ackerman et al.,2016, for a review), as well as in a social domain For example, calibration and resolution indices may lead to different interventions for social cognition. If overconfidence, measured by calibration, is high, then this may result in a lack of overall mental resource allocation to a situation (see Ackerman & Thompson, 2017). Therefore, if the goal is to increase mental resource allocation in social situations, such as re-evaluation or information seeking, then this self-monitoring process should be targeted. However, if discrimination between concepts that are more or less known, measured by resolution, is low, then this may result in an inefficient allocation of mental resources to skills or concepts that one already knows (see Ackerman & Thompson, 2017). Therefore, if the goal is to enable a person to use their mental resources efficiently in a social situation, for example by detecting and directing attention to social concepts that are less understood, then this self-monitoring process should be targeted.

Overall, by integrating measures of accuracy and confidence on an emotion recognition task, it was demonstrated that metacognitive abilities are measurable and relevant in a social domain.

Does Domain Matter? Domain-Generality versus Domain-Specificity

In the current study, I explored whether metacognitive monitoring represents a general ability, similar across domains, or rather a specific ability that varies as a function of the domain considered. Monitoring accuracy on the emotional face recognition task was compared to monitoring accuracy on the general knowledge task within individuals. As expected, calibration on the emotional face recognition task was positively correlated with calibration on the general knowledge task. This finding supports previous studies showing that calibration is domain-general, reflecting a stable individual difference trait (Jackson & Kleitman, 2014; 2016; Stanovich & West, 1997). In contrast, there was a lack of correlation between resolution across tasks. Discrimination between correct and incorrect choices on one task was not correlated with discrimination between correct and incorrect choices on the other task. This finding supports previous studies suggesting resolution represents a domain-specific ability that varies with domain content (Jackson et al., 2016).

The current findings suggest that calibration is consistent within individuals across areas. The consistency of calibration across diverse general knowledge and emotional face recognition tasks suggests that confidence may, in part, reflect a dispositional trait or tendency that is applied consistently across domains. This is consistent with previous research which support domain-generality (e.g., Erickson & Heit, 2015; Scott & Berman, 2013; Veenman and Verheij, 2001), and contrasts with previous research showing domain-specificity (e.g., Fitzgerald, Arvaneh, & Dockree, 2017; Glaser, 1991). If calibration represents a domain-general, dispositional tendency, then targeting calibration may lead to improved performance monitoring across domains.

In contrast to calibration, the current findings suggest that the same individual may have strong resolution on one task but low resolution on another. The inconsistency of resolution across general knowledge and emotional face recognition tasks suggests that the ability to discriminate between correct and incorrect answers in a social domain may be dependent on specific domain-content and knowledge, consistent with previous research showing domain-specific resolution (Jackson et al., , 2016). It is possible that individuals may display more analogous resolution abilities for tasks in which the underlying content knowledge is more closely related (e.g., algebra versus geometry), compared to tasks that rely on different knowledge bases (e.g., general knowledge versus emotion recognition). Given that previous studies in support of domain-general resolution have compared resolution across similar tasks (Ackerman & Beller, 2007; Finn & Metcalfe, 2008), findings from the present study suggest that when tasks are diverse, resolution might be domain-specific, rather than general.

The domain-generality of calibration and domain-specificity of resolution further reinforces previous claims that calibration and resolution capture different aspects of monitoring accuracy (Ais et al.,2016; Koriat et al. 2002; Koriat, 2012b; Maki et al., 2005; Thiede et al., 2015) and can yield important insights into interventions targeting metacognitive skill. For example, if the goal is to increase general metacognitive accuracy skills, overconfidence should be targeted, as this may apply across areas. If the goal is to increase metacognitive accuracy skills in a specific domain, discrimination skills relating to specific content knowledge should be targeted.

Taken together, the current study demonstrated that metacognitive monitoring may reflect some domain-general ability, but is dependent on the metacognitive metric used. Study findings support that calibration represents a person-centered, domain-general ability that is relatively consistent across areas. In contrast, resolution represents a domain-specific ability, sensitive to content knowledge. Such results reinforce previous claims that resolution and calibration are separable indicators of monitoring accuracy, both relevant for guiding an individual's effort regulation not only in a learning context as has previously been suggested, but also in a social context. A highly accurate individual would need to possess well-developed monitoring skills across both indices of calibration and resolution.

Individual Differences in Monitoring Accuracy

Another aim of the current study was to explore potential correlates and predictors of monitoring accuracy (see Figure 7 for significant predictors of monitoring accuracy).

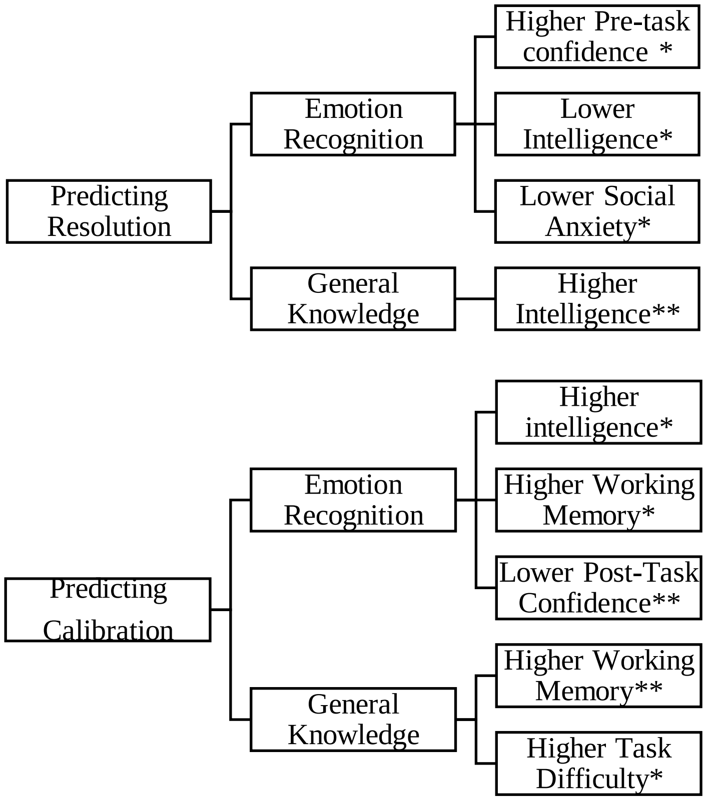


Figure 7. Significant predictors of calibration and resolution on the emotional face recognition and general knowledge tasks.

Cognitive Abilities Predictive of Monitoring Accuracy. Cognitive ability (either working memory or intelligence test scores) associated with calibration on the general knowledge and emotional face recognition tasks. Participants with greater cognitive ability were not only more successful on task items, but they were also better able to estimate their performance. These findings are consistent with other empirical studies that have examined overconfidence (Dentakos et al., 2019; Bruine de Bruine et al., 2007; Stanovich & West, 1998; Stanovich et al., 2016). Across tasks, cognitive ability predicted calibration. Calibration for emotional face recognition was predicted by intelligence and working memory. This finding may be attributed to mechanisms of cognitive decoupling. Cognitive decoupling is the ability to separate subjective information, while formally considering an alternative option (Stanovich, 2009). Cognitive decoupling may be important for supressing confidence bias on a task. For example, it may be important for a person to separate *feeling* confident in an answer based on how quickly that answer came to mind (separating subjective information), while considering alternative solutions. Cognitive decoupling can be challenging to achieve as decoupling mechanisms must be continuously ongoing and demand that mental simulations be sustained while keeping hypothetical scenarios decoupled (Stanovich, 2011). For these reasons, better decoupling abilities can be indexed with measures of working memory and intellectual abilities (Stanovich, 2011; West & Stanovich, 1997).

Cognitive ability (intelligence test scores) also correlated with resolution, but the direction of the relationship was task dependent. Cognitive ability was positively correlated with resolution for general knowledge, consistent with previous research (Jackson & Kleitman, 2014). In contrast, cognitive ability was negatively correlated with resolution for emotional face recognition. Behavioral regulation by emotions has been described as an autonomous process characterized by rapid, mandatory, execution when the triggering stimuli do not put a heavy load on processing capacity (Stanovich 2009; Stanovich et al. 2011). In general, continuous individual differences in autonomous processes are few, which may explain the findings with the emotion recognition task. Suppressing overconfidence may require some input form higher level control systems, but accurate discrimination of emotions may be more autonomous. It is more likely that cognitive abilities are unrelated than negatively related to resolution for emotion recognition. This is interpreted with caution, and replication will be critical to elucidate the relationship between monitoring accuracy and cognitive abilities in the domain of emotion recognition.

Verbal intelligence scores were correlated with confidence on the general knowledge task, but no cognitive ability measures correlated with confidence on the emotion recognition task. These findings further support that, as designed to be, tasks in the current study tapped into different knowledge areas, making cross-domain comparisons in monitoring accuracy possible.

Task-Specific Perceptions and Monitoring Accuracy. Another set of potential correlates explored in the current study were task-specific judgments, including predictive and postdictive aggregate confidence ratings and post-task ratings of difficulty, effort required, and feeling of effort.

Predictive confidence ratings were not correlated with accuracy for either task, suggesting that higher predictive assessments of performance were not matched with higher performance. However, *post*dictive confidence ratings positively correlated with accuracy for the general knowledge task. Therefore, lower accuracy on the task was matched with lower subjective assessments of overall performance after the task. In contrast, no relationship was found between accuracy and postdictive confidence ratings on the emotional face recognition task. These results suggest that previous findings positively linking aggregate judgments of performance to accuracy in various domains (Erikson & Heit, 2015) may not apply to the domain of emotion recognition. In an academic context, it has been proposed that aggregate confidence judgments are based on beliefs about one's abilities and previous experiences (Ackerman & Goldsmith, 2011; Koriat, 1997). This may similarly occur for aggregate confidence judgments for emotion recognition, but the lack of correlation between aggregate confidence and accuracy on the emotion recognition task suggests that aggregate judgments and the beliefs they tap into may not be reflective of objective performance. Aggregate confidence is a one-item measure. Self-efficacy is defined as one's beliefs about one's capabilities to produce designated levels of performance (Bandura, 1994). One way to explore how beliefs about one's abilities may impact one's performance monitoring in a social domain may be to examine self-efficacy, with a standardized measure, in relation to monitoring accuracy.

Postdictive confidence ratings negatively predicted calibration for emotion recognition. The better the calibration, the lower the confidence in aggregate performance. Postdictive confidence ratings did not predict calibration for general knowledge. Therefore, not only were higher postdictive judgments of performance not linked to higher accuracy for emotion recognition, they were also linked to poorer monitoring accuracy. Content in some domains may be more difficult to track for accuracy. Emotion recognition is found to be an automatic process (Ekman, 1972), and thus, people may have more difficulty tracking their accuracy and accessing metacognitive information in this domain than in an academic domain. Given the novelty of these findings, cautious interpretation and future work will be needed to elucidate these findings.

Predictive confidence negatively predicted resolution on the emotional face recognition task, suggesting that participants with lower predictive confidence have less discrimination between incorrect and correct emotion recognition. Predictive confidence did not predict resolution for the general knowledge task. Content in some domains, like emotion recognition, may be more “fuzzy” than discrete, which may impact discrimination. For example, determining the difference between the emotions of shyness and embarrassment may seem fuzzy compared to boundaries in general knowledge, where a capital city of a country is a concrete fact. It may be that individuals with lower predictive confidence, are more reluctant to discriminate between content that is more “fuzzy” (e.g., facial expressions) than concrete (e.g., general knowledge).

Confidence negatively correlated with perceptions of difficulty, effort, and discomfort across tasks. It is possible that items perceived as less difficult and effortful lead participants to have higher confidence ratings. However, the lack of correlation with monitoring accuracy indices for the emotional face recognition task suggests that participants did not rely strongly on perceived difficulty or effort to monitor performance. It may be that difficulty and effort ratings would impact monitoring accuracy for tasks that are experienced as extremely difficult or easy.

Theoretical Connections

The findings from the current study are consistent with a meta-reasoning framework proposed by Ackerman and Thompson (2017; see Figure 7). In this framework, meta-level processes govern cognitive processes (e.g., reasoning) by monitoring their performance (i.e., metacognitive monitoring; middle column of model) and allocating resources when required (i.e., metacognitive control; right column of model).

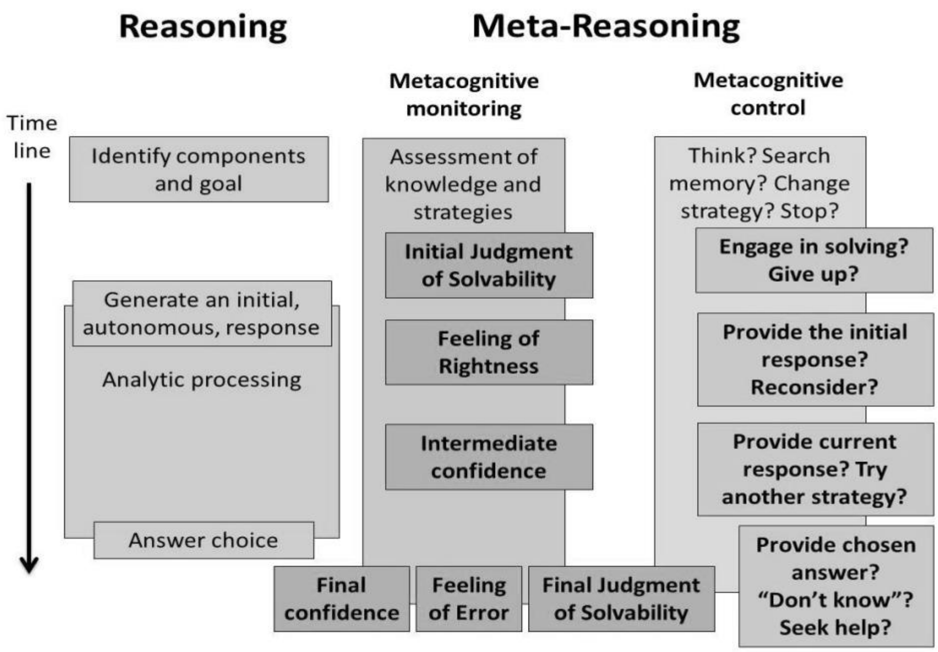


Figure 7. Proposed Meta-Reasoning Framework. Reprinted from Ackerman, R., & Thompson, V. A. (2017). Meta-reasoning: Monitoring and control of thinking and reasoning. *Trends in Cognitive Sciences*, *21*(8), 607-617.

Confidence ratings in the current study reflect the final confidence part of the framework. Final confidence is the subjective probability that a chosen answer is correct (Ackerman &Thompson, 2017). However, one limitation of relying on confidence ratings to empirically understand a person's metacognitive monitoring, is that confidence ratings alone do not provide a measure of monitoring accuracy. For example, a person may show high confidence, but this confidence may be warranted if this person is showing high performance. Final confidence may in reality guide the allocation of mental resources to a task, but using confidence ratings as a measure of monitoring accuracy renders it difficult to make comparisons between confidence and accuracy or to test predictions posited by the model. To test whether a person's confidence *appropriately* directs their investment of efforts, it is important to consider whether a person's confidence is or is not matched to accuracy.

The match or mismatch between item-specific confidence and accuracy is captured by indicators of calibration and resolution, which are not clearly defined in the Meta-Reasoning Framework (Ackerman & Thompson, 2017). Calibration and resolution capture the subjective probability that one's final response to a problem is correct *and* one's actual performance. In the current study, there was evidence to support that calibration and resolution represent different aspects of metacognitive monitoring, and this is not reflected in the Meta-reasoning Framework. Resolution is different from feeling of error (middle part of the model), which reflects the subjective feeling that an error was made (Ackerman & Thompson, 2017), as resolution refers to the ability to discriminate between correct and incorrect responses. One reason why differentiating monitoring accuracy indices within any framework for meta-reasoning is important is the differential effect that each index may have on subsequent control mechanisms. For example, overconfidence may result in a lack of investment in mental resources while poor discrimination may result in an inefficient allocation of resources to concepts one already understands. Including and differentiating between resolution and calibration as indicators of metacognitive monitoring in meta-reasoning frameworks may allow for a closer assessment of their respective influence on metacognitive control mechanisms.

Predictive and postdictive confidence judgments in the current study are similar to the initial and final judgments of solvability in the framework. However, predictive and postdictive judgments represent aggregate ratings, rather than item-specific ratings. Aggregate judgments are conceptually different from judgments made immediately before or after a single item is completed (Thiede et al., 2005; Veenman et al., 2006). An assessment of aggregate performance for a given domain taps into the cue-utilization view of metacognition (Koriat, 1997). It has been proposed that aggregative confidence judgments are formed based on beliefs about one's abilities and prior experiences (Ackerman & Goldsmith, 2011; Koriat, 1997). In the current study, it was demonstrated that predictive and postdictive ratings were not correlated with accuracy on the emotion recognition task, but they were correlated with confidence. These aggregate judgments may impact subsequent control processes (right column of model). For example, if overall knowledge in emotion recognition is judged as poor, an individual might seek more information (e.g., asking direct questions) the next time this individual encounters emotional expressions. In contrast, if overall knowledge is judged as sufficient, this behavior may not be engaged in. Therefore, aggregate judgments of knowledge may be linked not only to performance monitoring in a social domain, as shown the current study, but also in subsequent control mechanisms related to effort allocation and behavioural response. Thus, findings from the current study suggest that the Meta-Reasoning Framework also include aggregate judgments.

Overall, the main findings from the current study are consistent with the Meta-Reasoning Framework (Ackerman & Thompson, 2017), for monitoring accuracy more generally and in a social domain specifically, with some minor differences. It is suggested that the meta-reasoning framework include calibration and resolution indices and aggregate confidence ratings.

Monitoring Accuracy in a Social Domain and Self-Perceptions

Another goal of the current study was to explore self-perceptions as correlates of monitoring accuracy in a social domain.

Self-reported social anxiety, as measured by the Social Anxiety in University Students Questionnaire, did not correlate with success for either task, but self-reported social anxiety correlated with confidence for both tasks. This suggests that confidence, but not accuracy, varied with self-reported social anxiety. That is, the higher the social anxiety, the lower the confidence. Research has demonstrated that self-focused thoughts in individuals with social anxiety can lead to increased focus on negative self-perceptions (see Mor & Winquist, 2002; Clark and Wells, 1995; Hofmann, 2007; Perini, Abbott and Rapee, 2006). It may be that increased focus on negative self-perceptions produces a lowered confidence rating during self-monitoring.

Self-reported social anxiety predicted resolution on the emotional face recognition task, but the effect size was small. The lower the social anxiety scores, the better the discrimination between incorrect and correct emotion recognition. One mechanism that may explain why social anxiety scores predicted resolution for emotion recognition, and not general knowledge, is the biased recall of social events. Research has demonstrated a tendency for individuals with social anxiety to recall a social event as being more negative than it was (Hofmann, 2007; Brozvich & Heimberg, 2008). This may lead to difficulty separating accurate from inaccurate choices on a social task. Calibration was not correlated with social anxiety scores, consistent with research that showed no association between anxiety and monitoring accuracy in an academic domain (Jiang & Kleitman, 2014). In the current study, because social anxiety scores correlated with confidence but not success, participants with lower social anxiety can have high calibration (i.e., high confidence, high success) or low calibration (i.e., high confidence, low success).

Contrary to current hypotheses, perceived empathic ability as measured by the Toronto Empathy Questionnaire, did not correlate with calibration or resolution for emotional face recognition. These findings suggest that participants who perceive themselves to be more empathetic do not necessarily have more or less biased beliefs about the correctness of their performance in an emotion recognition domain. In addition, no correlation was found between self-reported emotion differentiation, as measured by the Differentiation of Emotions Scale, and monitoring accuracy indices. This finding suggests that participants with highly differentiated emotional states are no more likely than participants with poorly differentiated emotional states to assign the correct level of confidence to performance.

This is the first known study to explore self-perceptions in relation to monitoring accuracy in a social domain. Given the novelty and exploratory nature of the current findings, as well as the small effect size for the relationship between social anxiety and resolution, further research is needed to replicate and build upon the current findings.

Cognitive Emotion Regulation and Monitoring Accuracy. Participants rated their tendency to use cognitive coping skills that are adaptive (i.e., positive reappraisal; refocusing; planning) and maladaptive (i.e., self-blame; other-blame; catastrophizing) when faced with challenges. Contrary to the current hypotheses, increased endorsement of adaptive coping skills did not correlate with monitoring accuracy indices. The emotion regulation measure used in the current study is an index of self-reported use of coping skills. As there were overconfidence tendencies in the current sample (i.e., on experimental tasks), the endorsement of more adaptive coping skills may reflect participants' tendency to perceive themselves positively. In addition, the ability to regulate emotions may impact more on monitoring accuracy in scenarios that are experienced as highly stressful (e.g., aggressive conflicts). Future studies should include emotion regulation measures that incorporate "online" performance-based paradigms.

Monitoring Accuracy for Emotional Face Recognition and a Behavioural Correlate

A final goal in the current study was to explore monitoring accuracy on the emotional face recognition task in association with a behavioural correlate. Contrary to hypotheses, indices of monitoring accuracy were not correlated with self-reported social-emotional challenges. It is possible that the one-item measure of social-emotional changes in the current study, did not sufficiently capture behavioural challenges experienced by participants. In addition, perhaps monitoring accuracy in a social domain would be correlated with decision-making in social situations. Studies that have explored behavioural consequences of monitoring accuracy have focused on decision-making, such as whether to have a loan foreclosed (Bruine de Bruin et al., 2007). These studies have found that poorer monitoring accuracy leads to poorer decisions. Monitoring accuracy may be associated with decision-making in social situations, such as whether to use an addictive substance, seek advice, or trust a peer or doctor. If individuals poorly estimate their skills or comprehension or are unable to discriminate between situations in which they have mastered knowledge and those they have not, then they may be less likely to engage in behaviours that would favor the improvement of their skills.

Implications

Implications for Metacognitive Training. Studies have shown that metacognitive training can be beneficial for performance in an academic context, particularly for low-achieving groups (Cardell-Ellawar, 1995; Krugger & Denning, 1999; Kramarski & Mevarich, 2003; Teong, 2003; Bol et al., 2015). Similarly, metacognitive training may be beneficial for social behaviour in a community or clinical context, for groups with social impairments. If calibration and resolution represent distinct aspects of monitoring accuracy, then distinct intervention targets may be required to help support calibration and resolution.

Metacognitive Training of Calibration. Due to the domain-generality of calibration, increased training in facial emotions recognition, or other social skills, may not necessarily lead to better calibration. To improve calibration, individuals may generally need to learn how to monitor their performance during social situations without relying on their subjective feelings of confidence for guidance. The domain-generality of overconfidence suggests that general beliefs and cognitive dispositions should be targeted to improve calibration.

One way to improve calibration may be to target beliefs about the quality of one's knowledge. Some research has examined the effect of metacognitive training on schizophrenia symptomology (for a review, see Moritz and Woodword, 2014). Learning objectives for metacognitive training included increased awareness of cognitive biases, such as jumping to conclusions, and increased openness to changing one's beliefs. For example, participants are taught about the consequences of hasty decision-making and how to withhold strong judgments until sufficient evidence becomes available. Participants are also taught about the constructive nature of memory and encouraged to decrease confidence when evidence is lacking. Thus, to improve calibration, programs that aim to decrease hasty decision-making and increase fact checking may be helpful. To help participants tolerate negative feelings that may accompany decreased confidence, self-esteem may also be an important target in the training of calibration.

Another way to target calibration may be to foster the increased use of specific strategies. For example, metacognitive training of mathematical problem solving has included teaching learners how to self-address questions for comprehension (designed to encourage participants to reflect on a problem) and strategy (designed to encourage participants to consider which strategy to use) (Mevarech & Kramarski, 1997; Kramarski & Mevarech, 2003). Therefore, individuals with a tendency for overconfidence may be taught to more generally self-address questions that promote increased reflection before making a decision (e.g., “what do I know about this situation?”) and increased consideration of strategy use (e.g., “what tools can I use to help me understand this situation?”) across different domains.

Metacognitive Training of Resolution. There is little known research on ways to improve resolution through metacognitive training. In contrast to calibration, the domain-specificity of resolution abilities suggests that to better aid individuals in discriminating between correct and incorrect interpretations of facial expressions, increased exposure would be required.

One way to improve resolution may be through domain-specific practice. One study by Koriat and colleagues (2002) demonstrated that resolution abilities improved with repeated practice. As individuals accumulate greater content knowledge, they may be better able to detect when they do not know an answer and appropriately assign lower confidence to such problems. Therefore, increased exposure may be corrective for individuals, leading to better discrimination between social cues and concepts that are more or less known. In a study by Keith and Frese (2005), Error Management Training was shown to have a positive impact on self-regulated learning. In this training, learners are provided with complex learning materials and opportunities to explore solutions and make mistakes. Learners are helped to learn from mistakes, which theoretically promotes increased domain knowledge and error detection. This approach may similarly help in targeting resolution for social knowledge and problem-solving.

Overall, the broad training implication for metacognitive monitoring in a social domain is that resolution and calibration may require distinct interventions. Strategies that target general cognitive and behavioural approaches to decision-making may be effective at improving calibration while strategies that target error detection may be effective at improving resolution.

Implications for Clinical Research. A novel contribution of the current study is that it provides an empirical method for exploring how internal criteria, such as confidence judgments, can help individuals monitor and adjust their social inferences. The social domain examined in the current study, emotional face recognition, represents important aspects of social information processing and has been examined in various clinical groups with a focus on aptitude. Exploring social functioning from a metacognitive framework may provide a different account of why, beyond differences in actual social skill, some individuals are more socially competent than others. From the current findings, several implications emerged for clinical research.

The current study examined metacognitive monitoring in a non-clinical university sample, raising the question of whether findings would vary for a clinical sample. In the clinical literature, impairments in emotion recognition have been described in several psychological disorders (e.g., bipolar patients; Montag et al., 2010; autism spectrum disorder, ASD, Dziobek et al., 2006; narcissistic personality disorder; Ritter et al., 2011) and in sexual offenders (Hudson et al., 1993; Gery, Miljkovitch, Berthoz, & Soussignan, 2009). Metacognitive awareness of social impairments may be a critical step for successful social skill intervention (Verhoeven et al. 2012). If individuals are unaware of their social impairments, then effort may not be sufficiently allocated to a social situation, and further, to social skill intervention. Poor self-monitoring in a social context can have negative consequences. For example, if a person overestimates their ability to infer emotions, this person may miss an expression of sadness and an opportunity to provide social support. Increased self-monitoring may motivate increased acquisition and use of strategies. A few studies have compared confidence and accuracy on a task between clinical (e.g., BPD) and non-clinical groups, but studies are rare and not replicated. In addition, using resolution and calibration indices allows for an additional measure of discrimination between correct and incorrect answers (resolution index) and an integrated measure of confidence and accuracy (calibration index). Clinical research may benefit from using these metrics.

Another question to explore is whether acquiring metacognitive skills early in development may support social skill development. Metacognitive skills typically begin to develop in childhood (Veerman, 2005). Also typically developing in childhood is an increased awareness that mental states are internal, subjective experiences related to, but separate from, the external world (Wellman, 1990, 2011). However, not all children develop the ability to attribute mental states to others, which is proposed to play a critical role in social development (Denham, Blair, Schmidt, & DeMul-der, 2002; Astington, 2003; Lemerise & Arsenio, 2000). It is possible that early metacognitive training may help to facilitate increased engagement in early social skills intervention.

Limitations and Future Directions

Limitations.

Calculating indices. In the resolution literature, variability in item difficulty is important, but this screening is not used in the overconfidence literature. In order to calculate the Goodman-Kruskal Gamma index and whether a participant can discriminate between correct and incorrect items, there needs to be variability in confidence judgments (Fleming & Lau, 2014). When calculating calibration, this is less of a methodological concern because it is possible that a participant assigns high or low confidence ratings to all responses, due to a strong overconfidence or underconfidence bias. To my knowledge, this is the first study to directly and experimentally compare calibration and resolution indices to address issues of domain-generality versus specificity. In order to use the same set of responses and participants and to directly compare findings across accuracy monitoring indices, I eliminated items that had very low or very high success rates. This was considered to be the most conservative approach and was used in order to directly compare these indices. However, by eliminating items that had very low or very high success rates, it is possible that my understanding of monitoring accuracy when performance is extremely low or high is limited.

In addition, the Bias index is examined for statistical differences from zero to determine whether there is over- or underconfidence on a task (e.g., Bruine de Bruin et al.,2007; West & Stanovich, 1997). In the current study, there was significant overconfidence. However, the degree of overconfidence is not clearly interpretable using this methodology, nor is it possible to differentiate between varying degrees of overconfidence. An important consideration for future research is to consider the range of Bias index scores and which range of Bias scores would constitute varying degrees of low, moderate, or high overconfidence.

Item difficulty. Item difficulty is another consideration for future studies. Difficulty can impact confidence ratings, as easier questions tend to evoke underconfidence and harder questions tend to evoke overconfidence (Juslin et al., 2000). One of the challenges with exploring monitoring abilities across domains is designing experimental tasks from different domains that have items of similar difficulty. The tasks developed for this study were developed to be similar in difficulty. Continued consideration of item difficulty will be important in future studies. Furthermore, given the above-described methodological challenges of calculating calibration and resolution indices, consideration of similar item variability in cross-domain comparisons will be important.

Sample size. The sample size is a third limitation in the current study considering the number of completed analyses. Given the novelty of my research objectives and methodology, current hypotheses needed to be approached systematically.

Sample distribution. Another potential issue related to the sample in the current study is the predominantly female sex distribution. In the current study, I did not find any sex differences in monitoring accuracy abilities. However, this finding may have been impacted by the skewed female sample. Future studies should aim for a more equal distribution between males and females and further explore any potential sex differences.

Task reliability. Task reliability is another consideration for future research. Task reliability has generally not been reported in the overconfidence literature and item screening methods and variability of accuracy are most critical when examining resolution indices. However, in future studies, task reliability should be reported. It is acknowledged that higher reliability in the monitoring accuracy tasks could have impacted the results. Future studies can improve on issues of task reliability by including a larger set of task items and by increased piloting of task items.

A related potential limitation is one-item measures. The relative lack of findings with the task-specific ratings may be, partly, attributable to them being one-item measures, which increases risk for random error. However, task-specific ratings were based on the typical method used in the literature. Self-reported social-emotional challenges were also measured with one item, which may not have sufficiently captured social-emotional challenges experienced by participants. Future studies may include a standardized measure of social-emotional.

Future directions. Considering the novelty of this research, there are many ways to extend and refine the current findings. First, given the preliminary nature of the current findings on metacognitive monitoring in a social domain, cautious interpretation and further replication will be critical. Tasks should be standardized to ensure increased reliability. Second, described methodological concerns related to cross-domain and cross-index comparisons, such as considerations of item variability and task difficulty, should be empirically explored. An exploration of how task format (e.g., open-ended versus multiple choice questions) can impact performance monitoring is also warranted. Third, future research grounded in a developmental perspective may be useful to gain an understanding of how to best support the development of monitoring abilities. Given calibration represents a trait-like, domain-general ability, an in-depth understanding of overconfidence in children and whether overconfidence improves with early metacognitive training would be of value. Fourth, future research grounded in clinical contexts would also be of value, in order to better understand confidence biases and resolution in clinical populations characterized with social impairments. Both calibration and resolution indices should be considered in such studies. Moreover, monitoring accuracy across domains should also be considered in such studies to explore whether monitoring accuracy impairments are linked to specific content domain. Fifth, given a self-report measure of emotion regulation was used in the current study, a greater exploration of the impact of emotion regulation on monitoring accuracy may be warranted. Future studies may include an experimental paradigm that can directly assess for the impact of stress on monitoring accuracy. Sixth, future studies may examine the influence of motivation on monitoring accuracy. For example, studies may include different types of reinforcement for performance and assess for the impact on monitoring accuracy. Seventh, proposed implications of this study should be explored. For example, future studies should test how metacognitive monitoring relates to and impacts metacognitive regulation in a social domain across various types of tasks.

Conclusion

Metacognition has been conceptualized as self-directed performance monitoring processes that guide everyday decision-making (Flavel, 1979; Nelson & Narrens, 1990). Being viewed as relevant for self-directed learning within academic settings, much research has been done on metacognitive monitoring related to academic functioning. The primary purpose of the current study was to extend the study of metacognitive monitoring to a social domain, where people must make implicit judgments about their accuracy before deciding on an action. An important and novel contribution of the current study, is that calibration represented a general ability across diverse general knowledge and emotional face recognition domains, while resolution was specific to a given domain. The findings highlight the critical difference between monitoring indices. A focal implication of this study is that calibration and resolution represent different aspects of monitoring accuracy, both important for effective monitoring and subsequent effort regulation in emotional face recognition. From an applied perspective, different interventions may be required to target each index and across domains. To my knowledge, this is the first study to explore resolution across diverse domains and this distinction between both these levels of monitoring accuracy is important. Beyond theoretical and applied implications, my findings in a non-clinical sample of emerging adults also raise important questions for metacognitive research in a social domain across age groups and clinical populations. Examining performance and concurrent judgments of performance provide a novel avenue to study decision-making in a social context.

References

Abbott, M. J., & Rapee, R. M. (2004). Post-event rumination and negative self-appraisal in social phobia before and after treatment. *Journal of Abnormal Psychology*, *113*(1), 136.

Ackerman, R. (2019). Heuristic cues for meta-reasoning judgments: review and methodology. *Psychological Topics, 28*(1), 1-20.

Ackerman, R., & Beller, Y. (2017). Shared and distinct cue utilization for metacognitive judgments during reasoning and memorisation. *Thinking & Reasoning*, *23*(4), 376-408.

Ackerman, R., & Goldsmith, M. (2011). Metacognitive regulation of text learning: On screen versus on paper. *Journal of Experimental Psychology: Applied, 17*(1), 18-32

Ackerman, R., Parush, A., Nassar, F., & Shtub, A. (2016). Metacognition and system usability: Incorporating metacognitive research paradigm into usability testing. *Computers in Human Behavior*, *54*, 101-113.

Ackerman, R., & Thompson, V. A. (2017). Meta-reasoning: Monitoring and control of thinking and reasoning. *Trends in Cognitive Sciences*, *21*(8), 607-617.

Ackerman, R., & Zalmanov, H. (2012). The persistence of the fluency–confidence association in problem solving. *Psychonomic Bulletin & Review*, *19*(6), 1187-1192.

Ais, J., Zylberberg, A., Barttfeld, P., & Sigman, M. (2016). Individual consistency in the accuracy and distribution of confidence judgments. *Cognition*, *146*, 377-386.

Allemand, M., Steiger, A. E., & Fend, H. A. (2015). Empathy development in adolescence predicts social competencies in adulthood. *Journal of Personality*, *83*(2), 229-241.

Amin, N., Foa, E. B., & Coles, M. E. (1998). Negative interpretation bias in social phobia. *Behaviour research and therapy*, *36*(10), 945-957.

Anderson, M. C. M., & Thiede, K. W. (2008). Why do delayed summaries improve metacomprehension accuracy? *Acta Psychologica*, *128*(1), 110-118.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The “Reading the Mind in the Eyes” test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of child psychology and psychiatry*, *42*(2), 241-251.

Bhamani, S., & Hussain, N. (2012). Social anxiety in higher education learning context: Scale construction and reliability. *Golden Research Thoughts*, *2*(5).

Bol, L., Campbell, K. D., Perez, T., & Yen, C. J. (2016). The effects of self-regulated learning training on community college students’ metacognition and achievement in developmental math courses. *Community College Journal of Research and Practice*, *40*(6), 480-495.

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology*, *64*, 417-444.

Bruine de Bruin, W., Parker, A. M., & Fischhoff, B. (2007). Individual differences in adult decision-making competence. *Journal of Personality and Social Psychology*, *92*, 938-956.

Cardell-Elawar, M. (1995). Effects of metacognitive instruction on low achievers in mathematics problems. *Teaching and Teacher Education*, 11, 81-95.

Cartwright-Hatton, S., Tschernitz, N., & Gomersall, H. (2005). Social anxiety in children: social skills deficit, or cognitive distortion?. *Behaviour research and therapy*, *43*(1), 131-141.

Cassady, J. C., & Johnson, R. E. (2002). Cognitive test anxiety and academic performance. *Contemporary Educational Psychology*, *27*(2), 270-295.

Chen, P. P. (2002). Exploring the accuracy and predictability of the self-efficacy beliefs of seventh-grade mathematics students. *Learning and Individual Differences*, *14*(1), 77-90.

Clark, D. M., & McManus, F. (2002). Information processing in social phobia. *Biological psychiatry*, *51*(1), 92-100.

Clark, J. V., & Arkowitz, H. (1975). Social anxiety and self-evaluation of interpersonal performance. *Psychological reports*, *36*(1), 211-221.

Desender, K., Boldt, A., & Yeung, N. (2018). Subjective confidence predicts information seeking in decision making. *Psychological science*, *29*(5), 761-778.

Desoete, A., & Roeyers, H. (2006). Metacognitive macroevaluations in mathematical problem solving. *Learning and Instruction*, *16*(1), 12-25.Destan, N., & Roebers, C. M. (2015). What are the metacognitive costs of young children's overconfidence? *Metacognition and Learning*, *10*(3), 347-374.

Dodge, K. A., Price, J. M., Bachorowski, J. A., & Newman, J. P. (1990). Hostile attributional biases in severely aggressive adolescents. *Journal of abnormal psychology*, *99*(4), 385-392.

Dornic, S., Ekehammar, B., & Laaksonen, T. (1991). Tolerance for mental effort: Self-ratings related to perception, performance and personality. *Personality and Individual Differences, 12*(3), 313-319.

Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produces underachievement: Inaccurate self evaluations undermine students' learning and retention. *Learning and Instruction*, *22*(4), 271-280.

Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed self-assessment: Implications for health, education, and the workplace. *Psychological Science in the Public Interest*, *5*(3), 69-106. purchase. *De Economist*, *164*(1), 19-39.

Eisenberg, N., & Fabes, R. A. (1990). Empathy: Conceptualization, measurement, and relation to prosocial behavior. *Motivation and Emotion*, *14*(2), 131-149.

Eisenberg, N., Fabes, R. A., Guthrie, I. K., & Reiser, M. (2000). Dispositional emotionality and regulation: their role in predicting quality of social functioning. *Journal of personality and social psychology*, *78*(1), 136-157.

Ekman, P. (1992). An argument for basic emotions. *Cognition & emotion*, *6*(3-4), 169-200.

Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: a meta-analysis. *Psychological bulletin*, *128*(2), 203-235.

Ertekin, E., Koyuncu, A., Ertekin, B. A., & Özyıldırım, İ. (2015). Alexithymia in social anxiety disorder: is there a specific relationship or is it a feature of comorbid major depression?. *Anatolian Journal of Psychiatry*, *16*(2), 130-137.

Erickson, S., & Heit, E. (2015). Metacognition and confidence: Comparing math to other academic subjects. *Frontiers in Psychology*, *6*, 10.

Everson, H. T., Smodlaka, I., & Tobias, S. (1994). Exploring the relationship of test anxiety and metacognition on reading test performance: A cognitive analysis. *Anxiety, Stress & Coping: An International Journal*, *7*(1), 85-96.

Finn, B., & Metcalfe, J. (2008). Judgments of learning are influenced by memory for past test. *Journal of Memory and Language*, *58*(1), 19-34.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*, *34*(10), 906-911.

Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human Neuroscience*, *8*, 443-451.

Fitzgerald, L. M., Arvaneh, M., & Dockree, P. M. (2017). Domain-specific and domain-general processes underlying metacognitive judgments. *Consciousness and Cognition: An International Journal*, *49*, 264-277.

Fukaya, T. (2013). Explanation generation, not explanation expectancy, improves metacomprehension accuracy. *Metacognition and Learning*, *8*(1), 1-18.

Garnefski, N., & Kraaij, V. (2006). Cognitive emotion regulation questionnaire–development of a short 18-item version (CERQ-short). *Personality and Individual Differences*, *41*(6), 1045-1053.

Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L. M., & Woloshin, S. (2007). Helping doctors and patients make sense of health statistics. *Psychological Science in The Public Interest, 8*, 53-96.

Glaser, R. (1991). The maturing of the relationship between the science of learning and cognition and educational practice. *Learning and Instruction*, *1*(2), 129-144.

Grainger, C., Williams, D. M., & Lind, S. E. (2016). Metacognitive monitoring and control processes in children with autism spectrum disorder: Diminished judgment of confidence accuracy. *Consciousness and Cognition: An International Journal*, *42*, 65-74.

Gross, J. J. (1999). Emotion regulation: Past, present, future. *Cognition & Emotion*, *13*(5), 551-573.

Hacker, D. J., Bol, L., & Keener, M. C. (2008). Metacognition in education: A focus on calibration. In J. Dunlosky, & R. A. Bjork (Eds.), *Handbook of metamemory and memory; handbook of metamemory and memory* (pp. 429-455,) Psychology Press, New York, NY.

Han, S., Lerner, J. S., & Keltner, D. (2007). Feelings and consumer decision making: The appraisal-tendency framework. *Journal of consumer psychology*, *17*(3), 158-168.

Hoaken, P. N., Allaby, D. B., & Earle, J. (2007). Executive cognitive functioning and the recognition of facial expressions of emotion in incarcerated violent offenders, non-violent offenders, and controls. *Aggressive Behavior: Official Journal of the International Society for Research on Aggression*, *33*(5), 412-421.

Hsu, C. F., Eastwood, J. D., & Toplak, M. E. (2017). Differences in perceived mental effort required and discomfort during a working memory task between individuals at-risk and not at-risk for ADHD. *Frontiers in Psychology*, *8*, 407-415.

Hsu, C. F., Propp, L., Panetta, L., Martin, S., Dentakos, S., Toplak, M. E., & Eastwood, J. D. (2018). Mental effort and discomfort: Testing the peak-end effect during a cognitively demanding task. *PloS One*, *13*(2), e0191479.

Ickes, W. (1993). Empathic accuracy. *Journal of personality*, *61*(4), 587-610.

Ickes, W., Stinson, L., Bissonnette, V., & Garcia, S. (1990). Naturalistic social cognition: Empathic accuracy in mixed-sex dyads. *Journal of Personality and Social Psychology*, *59*(4), 730-742.

Jackson, S. A., & Kleitman, S. (2014). Individual differences in decision-making and confidence: capturing decision tendencies in a fictitious medical test. *Metacognition and Learning*, *9*(1), 25-49.

Jackson, S. A., Kleitman, S., Howie, P., & Stankov, L. (2016). Cognitive abilities, monitoring confidence, and control thresholds explain individual differences in heuristics and biases. *Frontiers in Psychology, 7*, 14.

Jacobse, A. E., & Harskamp, E. G. (2012). Towards efficient measurement of metacognition in mathematical problem solving. *Metacognition and Learning, 7*(2), 133-149.

Johnson, D. D., & Fowler, J. H. (2011). The evolution of overconfidence. *Nature*, *477*(7364), 317.

Juslin, P., Winman, A., & Olsson, H. (2000). Naive empiricism and dogmatism in confidence research: A critical examination of the hard–easy effect. *Psychological Review*, *107*(2), 384-396.

Kaur, H., & Kaur, S. (2015). Social Anxiety in Relation to Alexithymia Among Adolescents. *Journal of Psychosocial Research*, *10*(2), 421-429.

Kelemen, W. L., Frost, P. J., & Weaver, C. A. (2000). Individual differences in metacognition: Evidence against a general metacognitive ability. *Memory & Cognition*, *28*(1), 92-107.

Kelly, K. J., & Metcalfe, J. (2011). Metacognition of emotional face recognition. *Emotion*, *11*(4), 896.

Keren, G. (1991). Calibration and probability judgments: Conceptual and methodological issues. *Acta Psychologica*, *77*(3), 217-273.

Klayman, J., Soll, J. B., Gonzalez-Vallejo, C., & Barlas, S. (1999). Overconfidence: It depends on how, what, and whom you ask. *Organizational Behavior and Human Decision Processes*, *79*(3), 216-247.

Kleitman, S. (2008). *Metacognition in the Rationality Debate. Self-confidence and its Calibration*. Saarbrucken, Germany: VDM Verlag Dr Muller.

Kleitman, S., & Stankov, L. (2001). Ecological and person-oriented aspects of metacognitive processes in test-taking. *Applied Cognitive Psychology, 15*(3), 321-341.

Kleitman, S., & Stankov, L. (2007). Self-confidence and metacognitive processes. *Learning and Individual Differences, 17*, 161-173.

Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.) Guilford Press, New York, NY.

Kohler, C. G., Turner, T. H., Bilker, W. B., Brensinger, C. M., Siegel, S. J., Kanes, S. J., ... & Gur, R. C. (2003). Facial emotion recognition in schizophrenia: intensity effects and error pattern. *American Journal of Psychiatry*, *160*(10), 1768-1774.

Komori, M. (2016). Effects of working memory capacity on metacognitive monitoring: A study of group differences using a listening span test. *Frontiers in Psychology, 7*, 285.

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126*(4), 349-370.

Koriat, A. (2008). Easy comes, easy goes? The link between learning and remembering and its exploitation in metacognition. *Memory & Cognition, 36*(2), 416-428.

Koriat, A. (2012a). The relationships between monitoring, regulation and performance. *Learning and Instruction*, *22*(4), 296-298.

Koriat, A. (2012b). The self-consistency model of subjective confidence. *Psychological Review, 119*(1), 80-113.

Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory, 6*(2), 107-118.

Koriat, A., & Shitzer-Reichert, R. (2002). Metacognitive judgments and their accuracy: Insights from the processes underlying judgments of learning in children. In M. Izaute, P. Chambres, & P.-J. Marescaux (Eds.), *Metacognition: Process, function, and use* (pp. 1–17). New York: Kluwer.

Koriat, A., Sheffer, L., & Ma'ayan, H. (2002). Comparing objective and subjective learning curves: judgments of learning exhibit increased underconfidence with practice. *Journal of Experimental Psychology: General*, *131*(2), 147-162.

Kramarski, B., & Mevarech, Z. R. (2003). Enhancing mathematical reasoning in the classroom: The effects of cooperative learning and metacognitive training. *American Educational Research Journal*, *40*(1), 281-310.

Kröner, S., & Biermann, A. (2007). The relationship between confidence and self-concept - towards a model of response confidence. *Intelligence, 35*(6), 580-590.

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), 1121-1134.

Lauterman, T., & Ackerman, R. (2014). Overcoming screen inferiority in learning and calibration. *Computers in Human Behavior*, *35*, 455-463.

Lee, J. (2009). Universals and specifics of math self-concept, math self-efficacy, and math anxiety across 41 PISA 2003 participating countries. *Learning and Individual Differences, 19*(3), 355-365.

Legg, A. M., & Locker, L., Jr. (2009). Math performance and its relationship to math anxiety and metacognition. *North American Journal of Psychology, 11*(3), 471-486.

Lemerise, E. A., & Arsenio, W. F. (2000). An integrated model of emotion processes and cognition in social information processing. *Child development*, *71*(1), 107-118.

Lennarz, H. K., Lichtwarck-Aschoff, A., Timmerman, M. E., & Granic, I. (2018). Emotion differentiation and its relation with emotional well-being in adolescents. *Cognition and Emotion, 32*(3), 651-657.

Lerner, J. S., & Keltner, D. (2000). Beyond valence: Toward a model of emotion-specific influences on judgment and choice. *Cognition & emotion*, *14*(4), 473-493.

Lichtenstein, S., & Fischhoff, B. (1977). Do those who know more also know more about how much they know?. *Organizational Behavior and Human Performance*, *20*(2), 159-183.

Lockl, K., & Schneider, W. (2002). Developmental trends in children's feeling-of-knowing judgments. *International Journal of Behavioral Development*, *26*(4), 327-333

Lusardi, A., Mitchell, O. S., & Curto, V. (2009). *Financial literacy among the young: Evidence and implications for consumer policy*. St. Louis: Federal Reserve Bank of St Louis.

Lusardi A., & Tufano, P. (2015). Debt literacy, financial experiences, and overindebtedness. *Journal of Pension Economics & Finance, 14*(4), 332-368.

Macher, D., Papousek, I., Ruggeri, K., & Paechter, M. (2015). Statistics anxiety and performance: Blessings in disguise. *Frontiers in Psychology, 6,* 4.

Maki, R. H., Shields, M., Wheeler, A. E., & Zacchilli, T. L. (2005). Individual Differences in Absolute and Relative Metacomprehension Accuracy. *Journal of Educational Psychology*, *97*(4), 723-731.

Malmendier, U., & Tate, G. (2008). Who makes acquisitions? CEO overconfidence and the market's reaction. *Journal of Financial Economics, 89*(1), 20-43.

Martin, R. C., & Dahlen, E. R. (2005). Cognitive emotion regulation in the prediction of depression, anxiety, stress, and anger. *Personality and individual differences*, *39*(7), 1249-1260.

Masson, M. E., & Rotello, C. M. (2009). Sources of bias in the Goodman–Kruskal gamma coefficient measure of association: Implications for studies of metacognitive processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(2), 509-527.

McAlpine, C., Kendall, K. A., & Singh, N. N. (1991). Recognition of facial expressions of emotion by persons with mental retardation. *American Journal on Mental Retardation, 96*(1), 29-36.

Metcalfe, J., & Shimamura, A. P. (1995). Metacognition: Knowing about knowing. *Information Processing and Management*, *31*(2), 261-262.

Miller, T. M., & Geraci, L. (2011). Unskilled but aware: reinterpreting overconfidence in low-performing students. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(2), 502-506.

Miron-Shatz, T., & Hanoch, Y., Doniger, G., Omer, B., & Ozanne, E. (2014). Subjective but not objective numeracy influences willingness to pay for BRCA1/2 genetic testing. *Judgment and Decision Making. 9.* 152-158.

Montagne, B., Schutters, S., Westenberg, H. G., van Honk, J., Kessels, R. P., & de Haan, E. H. (2006). Reduced sensitivity in the recognition of anger and disgust in social anxiety disorder. *Cognitive Neuropsychiatry*, *11*(4), 389-401.

Moritz, S., Andreou, C., Schneider, B. C., Wittekind, C. E., Menon, M., Balzan, R. P., & Woodward, T. S. (2014). Sowing the seeds of doubt: a narrative review on metacognitive training in schizophrenia. *Clinical psychology review*, *34*(4), 358-366.

Morrison, A. S., Mateen, M. A., Brozovich, F. A., Zaki, J., Goldin, P. R., Heimberg, R. G., & Gross, J. J. (2016). Empathy for positive and negative emotions in social anxiety disorder. *Behaviour research and therapy*, *87*, 232-242.

Morsanyi, K., Cheallaigh, N. N., & Ackerman, R. (2019). Mathematics anxiety and metacognitive processes: proposal for a new line of inquiry. *Psychological Topics*, *28*(1), 147-169.

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin*, *95*(1), 109-133.

Nelson, T. O., & Narens, L. (1980). Norms of 300 general-information questions: Accuracy of recall, latency of recall, and feeling-of-knowing ratings. *Journal of Verbal Learning and Verbal Behavior*, *19*(3), 338-368.

Özsoy, G., & Ataman, A. (2017). The effect of metacognitive strategy training on mathematical problem solving achievement. *International Electronic Journal of Elementary Education*, *1*(2), 67-82.

Pallier, G., Wilkinson, R., Danthir, V., Kleitman, S., Knezevic, G., Stankov, L., & Roberts, R. D. (2002). The role of individual differences in the accuracy of confidence judgments. *Journal of General Psychology, 129*(3), 257-299.

Parker, A. M., & Fischhoff, B. (2005). Decision-making competence: External validation through an individual differences approach. *Journal of Behavioral Decision Making, 18*, 1-27.

Perfect, T. J. (2004). The role of self-rated ability in the accuracy of confidence judgments in eyewitness memory and general knowledge. *Applied Cognitive Psychology, 18*(2), 157-168.

Phillips, L. H., Channon, S., Tunstall, M., Hedenstrom, A., & Lyons, K. (2008). The role of working memory in decoding emotions. *Emotion, 8*(2), 184-191.

Pressley, M. (2002). Metacognition and self-regulated comprehension. In A. E. Farstrup & S. J. Samuels, (Eds.), What research has to say about reading instruction (3rd ed., pp. 291 -309). Newark, DE: International Reading Association.

Pugalee, D. K. (2001). Writing, mathematics, and metacognition: Looking for connections through students' work in mathematical problem solving. *School Science and Mathematics, 101*(5), 236-245.

Reynolds, W. M., Ramírez, M. P., Magriña, A., & Allen, J. E. (1980). Initial development and validation of the academic self-concept scale. *Educational and Psychological Measurement, 40*(4), 1013-1016.

Rinne, L. F., & Mazzocco, M. M. M. (2014). Knowing right from wrong in mental arithmetic judgments: Calibration of confidence predicts the development of accuracy. *PLoS One, 9*(7).

Ronis, D. L., & Yates, J. F. (1987). Components of probability judgment accuracy: Individual consistency and effects of subject matter and assessment method. *Organizational Behavior and Human Decision Processes, 40*(2), 193-218.

Rozenblit, L., & Keil, F. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science, 26*(5), 521-562.

Schneider, W., Visé, M., Lockl, K., & Nelson, T. O. (2000). Developmental trends in children's memory monitoring: Evidence from a judgment-of-learning task. *Cognitive Development, 15*(2), 115-134.

Schrand, C. M., & Zechman, S. L. (2012). Executive overconfidence and the slippery slope to financial misreporting. *Journal of Accounting and Economics*, *53*(1-2), 311-329.

Schraw, G. (2009). A conceptual analysis of five measures of metacognitive monitoring. *Metacognition and learning*, *4*(1), 33-45.

Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology, 19*(4), 460-475.

Schraw, G., Dunkle, M. E., Bendixen, L. D., & Roedel, T. D. (1995). Does a general monitoring skill exist?. *Journal of Educational Psychology, 87(3), 433-444.*

Scott, B. M., & Berman, A. F. (2013). Examining the domain-specificity of metacognition using academic domains and task-specific individual differences. *Australian Journal of Educational & Developmental Psychology, 13*, 28-43.

Shipley, W. C., Gruber, C. P., Martin, T. A., & Klein, A. M. (2009). Shipley-2. *Los Angeles, CA: Western Psychological Services*.

Sidi, Y., Shpigelman, M., Zalmanov, H., & Ackerman, R. (2017). Understanding metacognitive inferiority on screen by exposing cues for depth of processing. *Learning and Instruction, 51*, 61-73.

Simon, E. W., Rosen, M., Grossman, E., & Pratowski, E. (1995). The relationships among emotional face recognition, social skills, and quality of life. *Research in Developmental Disabilities*, *16*(5), 383-391.

Simonian, S. J., Beidel, D. C., Turner, S. M., Berkes, J. L., & Long, J. H. (2001). Recognition of facial affect by children and adolescents diagnosed with social phobia. *Child Psychiatry and human development*, *32*(2), 137-145.

Soderstrom, N. C., Yue, C. L., & Bjork, E. L. (2015). Metamemory and education. In *The Oxford Handbook of Metamemory*.

Sporer, S. L., Penrod, S., Read, D., & Cutler, B. (1995). Choosing, confidence, and accuracy: A meta-analysis of the confidence-accuracy relation in eyewitness identification studies. *Psychological Bulletin*, *118*(3), 315, 327.

Stankov, L., Kleitman, S., & Jackson, S. A. (2014). Measures of the trait of confidence. In G. J. Boyle, H. Saklofske, & G. Matthews (Eds.), *Measures of personality and social psychological constructs*. Academic Press (pp. 158-189).

Stankov, L., Lee, J., Luo, W., & Hogan, D. J. (2012). Confidence: A better predictor of academic achievement than self-efficacy, self-concept and anxiety? Learning and Individual Differences, 22(6), 747-758.

Stanovich, K. (2011). *Rationality and the reflective mind*. New York, NY: Oxford University Press.

Stanovich, K. E. (2009). *What intelligence tests miss: The psychology of rational thought*. New Haven, CT: Yale University Press.

Stanovich, K. E., & West, R. F. (1998). Individual differences in rational thought. *Journal of Experimental Psychology: General*, *127*(2), 161-188.

Stanovich, K. E., West, R. F., & Toplak, M. E. (2016). *The Rationality Quotient: Toward a test of rational thinking*. MIT Press.

Thiede, K. W., Anderson, M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology*, *95*(1), 66-73.

Thiede, K. W., Dunlosky, J., Griffin, T. D., & Wiley, J. (2005). Understanding the delayed-keyword effect on metacomprehension accuracy. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(6), 1267-1280.

Thiede, K. W., Griffin, T. D., Wiley, J., & Anderson, M. C. (2010). Poor metacomprehension accuracy as a result of inappropriate cue use. *Discourse Processes*, *47*(4), 331-362.

Thiede, KW, Muller, ML, & Dunlosky, J. (2015) Methodology for investigating human metamemory: Problems and pitfalls: The Oxford handbook of metamemory.

Thiede, K. W., Wiley, J., & Griffin, T. D. (2011). Test expectancy affects metacomprehension accuracy. *British Journal of Educational Psychology*, *81*(2), 264-273.

Thomas, L. A., De Bellis, M. D., Graham, R., & LaBar, K. S. (2007). Development of emotional facial recognition in late childhood and adolescence. *Developmental science*, *10*(5), 547-558.

Thompson, V. A. (2009). Dual process theories: A metacognitive perspective. In J. ST. B. T. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond. Oxford University Press, Oxford*.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of memory and language*, *28*(2), 127-154.

Tauber, S. K., Dunlosky, J., Rawson, K. A., Rhodes, M. G., & Sitzman, D. M. (2013). General knowledge norms: Updated and expanded from the Nelson and Narens (1980) norms. *Behavior Research Methods*, *45*(4), 1115-1143.

Vasiliki Orgeta & Louise H. Phillips (2007) Effects of age and emotional intensity on the recognition of facial emotion, experimental aging research, 34:1, 63-79, DOI:10.1080/03610730701762047

Veenman, M. V. J., & Spaans, M. A. (2005). Relation between intellectual and metacognitive skills: Age and task differences. *Learning and Individual Differences, 15*(2), 159-176.

Veenman, M. V., Van Hout-Wolters, B. H., & Afflerbach, P. (2006). Metacognition and learning: Conceptual and methodological considerations. *Metacognition and Learning*, *1*(1), 3-14.

Veenman, M. V. J., & Verheij, J. (2003). Technical students' metacognitive skills: Relating general vs. specific metacognitive skills to study success. *Learning and Individual Differences, 13*(3), 259-272.

West, R. F., & Stanovich, K. E. (1997). The domain specificity and generality of overconfidence: Individual differences in performance estimation bias. *Psychonomic Bulletin & Review*, *4*(3), 387-392.

Wilson, T. D., & Brekke, N. (1994). Mental contamination and mental correction: unwanted influences on judgments and evaluations. *Psychological bulletin*, *116*(1), 117-142.

Wong, N., Beidel, D. C., Sarver, D. E., & Sims, V. (2012). Emotional face recognition in children with high functioning autism and children with social phobia. *Child Psychiatry & Human Development*, *43*(5), 775-794.

Yang, J. W., Yoon, K. L., Chong, S. C., & Oh, K. J. (2013). Accurate but pathological: Social anxiety and ensemble coding of emotion. *Cognitive Therapy and Research*, *37*(3), 572-578.

Yates, J. F. (1990). *Judgment and decision making* Prentice-Hall, Inc, Englewood Cliffs, NJ.

Yates, J. F., Lee, J., & Bush, J. G. (1997). General knowledge overconfidence: Cross-national variations, response style, and "reality". *Organizational Behavior and Human Decision Processes*, *70*, 87-94.

Zabrucky, K. M. (2010). Knowing what we know and do not know: Educational and real-world implications. *Procedia-Social and Behavioral Sciences*, *2*(2), 1266-1269.

Zacharakis, A. L., & Shepherd, D. A. (2001). The nature of information and overconfidence on venture capitalists' decision making. Journal of Business Venturing, 16(4), 311-332.

APPENDIX A: MEASURES

Demographic Information

Age: ________

Gender:
___ Male
___ Female
___ Other

_Year in University:
___ 1st year undergrad
___ 2nd year undergrad
___ 3rd year undergrad
___ 4th year undergrad
___ 5th year undergrad
___ Post-BA Continuing
___ Other

Please indicate your ethnicity (Check one)
___White/European
___ Black
___Asian
___Aboriginal
___South-Asian
___Arab
___ Latino-Hispanic
___ Other (please specify)________________

Is English your first language? _Yes _ No
If No, how long have you been speaking English? ____________ Years

Financially, do you consider your family to be:
___Well below average income
___Below average income
___Average income
___Above average income
___Well above average income

Current academic average grade:
___Below 49%
___50 – 59%
___60 – 69%
___70 – 79%
___80 – 100%

Overall, do you think that you have difficulties in one or more of the following areas: emotions, behavior or being able to get along with other people?
____No
____Yes – minor difficulties
____Yes – definite difficulties
____Yes – severe difficulties

Overall, do you think that you have difficulties in learning or academics?
____No
____Yes – minor difficulties
____Yes – definite difficulties
____Yes – severe difficulties

Pease check the column that indicates the highest level of education for each parent**:**

	Mother	**Father**
Less than 7th Grade	☐	☐
Junior high / Middle school (9th grade)	☐	☐
Partial high school (10th or 11th grade)	☐	☐
High school graduate	☐	☐
Partial college/university (at least one year)	☐	☐
College/university education	☐	☐
Graduate/professional degree	☐	☐

APPENDIX B: EXPERIMENTAL TASKS PILOT DATA

General Knowledge Test

Item	Correct Response	Success Rate (%)
1. What is the longest river in South America?	Amazon	63.16
2. For which country is the yen the monetary unit?	Japan	57.89
3. What is the last name of the man who first studied genetic inheritance in plants?	Mendel	52.63
4. What is the proper name for a badminton bird?	Shuttlecock	36.84
5. What is the last name of the author who wrote “Oliver Twist”?	Dickens	63.16
6. What is the last name of the man who invented the phonograph?	Edison	33.33
7. What is the name of an inability to sleep?	Insomnia	94.74
8. What is the name of the lizard that changes its colour to match the surroundings?	Chameleon	100.00
9. What is the name of the largest ocean on earth?	Pacific	36.84
10. What is the capital of Australia?	Canberra	26.32
11. What animal runs the fastest?	Cheetah	100
12. What is the term for hitting a volleyball down hard into the opponents court?	Spike	89.47
13. What is the name of the brightest star in the sky excluding the sun?	Sirius	52.63
14. What is the name of a dried grape?	Raisin	84.21
15. What is the name of the largest desert on earth?	Antarctica	5.26
16. What is the name of the mountain range that separates Asia from Europe?	Ural	15.79
17. What is the name for a medical doctor who specializes in diseases of the skin?	Dermatologist	47.37
18. What is the unit of sound intensity?	Decibel	63.16
19. What is the name of deer meat?	Venison	57.89
20. What is the name of the organ that produces insulin?	Pancreas	68.42
21. What is the name of the automobile instrument that measures mileage?	Odometer	63.16
22. What is the name of the bird that cannot fly and is the largest bird on earth?	Ostrich	94.74
23. What is the name for a cyclone that occurs over land?	Tornado	78.95
24. What is the largest plan in the solar system?	Jupiter	63.16

Note. N = 18; total success rate = 60.38

Emotion Recognition Test

Item	Correct Response	Valence	Success Rate (%)
1	Playful	Positive	83.33
2	Upset	Negative	66.67
3	Desire	Positive	61.11
4	Insisting	Ambiguous	61.11
5	Worried	Negative	77.78
6	Uneasy	Negative	41.18
7	Despondent	Negative	88.89
8	Regretful	Negative	66.67
9	Anticipating	Ambiguous	83.33
10	Contemplative	Ambiguous	50.00
11	Thoughtful	Ambiguous	61.11
12	Doubtful	Ambiguous	70.59
13	Decisive	Positive	77.78
14	Tentative	Negative	61.11
15**	Friendly	Positive	88.89
16	Interested	Positive	77.78
17	Cautious	Ambiguous	83.33
18	Interested	Positive	83.33
19	Reflective	Ambiguous	47.06
20	Flirtatious	Positive	82.35
21**	Confident	Positive	58.82
22	Serious	Ambiguous	64.70
23	Nervous	Negative	68.75
24	Suspicious	Negative	76.47

Note. *N* = 18; total success rate = 69.00

APPENDIX C: EXPERIMENTAL TASKS STUDY DATA

General Knowledge Test

Item	Correct Response	Success Rate (%)	Mean Confidence (%)
1. What is the longest river in South America?	Amazon	54.4	61
2. For which country is the yen the monetary unit?	Japan	37.5	71
3. What is the last name of the man who first studied genetic inheritance in plants?	Mendel	59.6	71
4. What is the proper name for a badminton bird?	Shuttlecock	34.6	81
5. What is the last name of the author who wrote "Oliver Twist"?	Dickens	46.3	54
6. What is the last name of the man who invented the phonograph?	Edison	25.7	57
7. What is the name of an inability to sleep?[a]	Insomnia	91.2	92
8. What is the name of the lizard that changes its colour to match the surroundings? [a]	Chameleon	96.3	94
9. What is the name of the largest ocean on earth?	Pacific	58.8	76
10. What is the capital of Australia? [a]	Canberra	9.6	77
11. What animal runs the fastest? [a]	Cheetah	94.9	92
12. What is the term for hitting a volleyball down hard into the opponents court? [a]	Spike	93.4	88
13. What is the name of the brightest star in the sky excluding the sun? **	Sirius	19.4	68
14. What is the name of a dried grape? **	Raisin	87.8	91
15. What is the name of the largest desert on earth? [a]	Antarctica	3.7	80
16. What is the name of the mountain range that separates Asia from Europe? [a]	Ural	6.6	63
17. What is the name for a medical doctor who specializes in diseases of the skin? [a]	Dermatolog ist	94.9	94
18. What is the unit of sound intensity?	Decibel	54.4	79
19. What is the name of deer meat?	Venison	47.1	61
20. What is the name of the organ that produces insulin?	Pancreas	68.4	73
21. What is the name of the automobile instrument that measures mileage?	Odometer	40.4	76
22. What is the name of the bird that cannot fly and is the largest bird on earth?	Ostrich	86.0	88
23. What is the name for a cyclone that occurs over land?	Tornado	75.0	79
24. What is the largest plan in the solar system?	Jupiter	51.5	74

Note. N = 136; ** *N* = 98; Total success rate = 53.19%

[a]Item removed due to overall success of less than 10% or greater than 90%.

Emotional Face Recognition Test

Item	Correct Answer	Valence	Success (%)	Mean Confidence (%)
1	Playful	Positive	57.36	77.76
2	Upset	Negative	66.91	81.62
3	Desire	Positive	67.65	76.10
4	Insisting	Ambiguous	59.56	74.37
5	Worried	Negative	78.67	83.16
6	Uneasy	Negative	57.35	73.16
7	Despondent	Negative	86.77	75.46
8	Regretful	Negative	74.26	74.32
9	Anticipating	Ambiguous	77.94	75.38
10	Contemplative	Ambiguous	65.44	74.72
11	Thoughtful	Ambiguous	66.12	78.73
12	Doubtful	Ambiguous	59.56	73.61
13	Decisive	Positive	55.88	73.59
14	Tentative	Negative	53.67	71.54
15**	Friendly	Positive	78.35	78.87
16	Interested	Positive	58.82	74.46
17	Cautious	Ambiguous	72.64	75.91
18	Interested	Positive	63.26	74.03
19	Reflective	Ambiguous	43.38	76.77
20	Flirtatious	Positive	58.08	77.59
21**	Confident	Positive	55.67	75.82
22	Serious	Ambiguous	70.58	77.42
23	Nervous	Negative	67.65	77.52
24[a]	Suspicious	Negative	90.44	83.49

Note. *N* = 136; ** *N* = 98; Total accuracy = 65.94%

[a]Item removed due to overall success of greater than 90%.

APPENDIX D: GLOSSARY OF TERMS FOR THE EMOTIONAL FACE RECOGNITION TASK

ACCUSING	blaming The policeman was accusing the man of stealing a wallet.
AFFECTIONATE	showing fondness towards someone Most mothers are affectionate to their babies by giving them lots of kisses and cuddles.
AGHAST	horrified, astonished, alarmed Jane was aghast when she discovered her house had been burgled.
ALARMED	fearful, worried, filled with anxiety Claire was alarmed when she thought she was being followed home.
AMUSED	finding something funny I was amused by a funny joke someone told me.
ANNOYED	irritated, displeased Jack was annoyed when he found out he had missed the last bus home.
ANTICIPATING	expecting At the start of the football match, the fans were anticipating a quick goal.
ANXIOUS	worried, tense, uneasy The student was feeling anxious before taking her final exams.
APOLOGETIC	feeling sorry The waiter was very apologetic when he spilt soup all over the customer.
ARROGANT	conceited, self-important, having a big opinion of oneself The arrogant man thought he knew more about politics than everyone else in the room.
ASHAMED	overcome with shame or guilt The boy felt ashamed when his mother discovered him stealing money from her purse.

ASSERTIVE	confident, dominant, sure of oneself The assertive woman demanded that the shop give her a refund.
BAFFLED	confused, puzzled, dumbfounded The detectives were completely baffled by the murder case.
BEWILDERED	utterly confused, puzzled, dazed The child was bewildered when visiting the big city for the first time.
CAUTIOUS	careful, wary Sarah was always a bit cautious when talking to someone she did not know.
COMFORTING	consoling, compassionate The nurse was comforting the wounded soldier.
CONCERNED	worried, troubled The doctor was concerned when his patient took a turn for the worse.
CONFIDENT	self-assured, believing in oneself The tennis player was feeling very confident about winning his match.
CONFUSED	puzzled, perplexed Lizzie was so confused by the directions given to her, she got lost.
CONTEMPLATIVE	reflective, thoughtful, considering John was in a contemplative mood on the eve of his 60th birthday.
CONTENTED	satisfied After a nice walk and a good meal, David felt very contented.
CONVINCED	certain, absolutely positive Richard was convinced he had come to the right decision.
CURIOUS	inquisitive, inquiring, prying Louise was curious about the strange shaped parcel.
DECIDING	making your mind up The man was deciding whom to vote for in the election.

DECISIVE	already made your mind up Jane looked very decisive as she walked into the polling station.
DEFIANT	insolent, bold, don’t care what anyone else thinks The animal protester remained defiant even after being sent to prison.
DEPRESSED	miserable George was depressed when he didn't receive any birthday cards.
DESIRE	passion, lust, longing for Kate had a strong desire for chocolate.
DESPONDENT	gloomy, despairing, without hope Gary was despondent when he did not get the job he wanted.
DISAPPOINTED	displeased, disgruntled Manchester United fans were disappointed not to win the Championship.
DISPIRITED	glum, miserable, low Adam was dispirited when he failed his exams.
DISTRUSTFUL	suspicious, doubtful, wary The old woman was distrustful of the stranger at her door.
DOMINANT	commanding, bossy The sergeant major looked dominant as he inspected the new recruits.
DOUBTFUL	dubious, suspicious, not really believing Mary was doubtful that her son was telling the truth.
DUBIOUS	doubtful, suspicious Peter was dubious when offered a surprisingly cheap television in a pub.
EAGER	keen On Christmas morning, the children were eager to open their presents.
EARNEST	having a serious intention

	Harry was very earnest about his religious beliefs.
EMBARRASSED	ashamed After forgetting a colleague's name, Jenny felt very embarrassed.
ENCOURAGING	hopeful, heartening, supporting All the parents were encouraging their children in the school sports day.
ENTERTAINED	absorbed and amused or pleased by something I was very entertained by the magician.
ENTHUSAISTIC	very eager, keen Susan felt very enthusiastic about her new fitness plan.
FANTASIZING	daydreaming Emma was fantasizing about being a film star.
FASCINATED	captivated, really interested At the seaside, the children were fascinated by the creatures in the rock pools.
FEARFUL	terrified, worried In the dark streets, the women felt fearful.
FLIRTATIOUS	brazen, saucy, teasing, playful Connie was accused of being flirtatious when she winked at a stranger at a party.
FLUSTERED	confused, nervous and upset Sarah felt a bit flustered when she realized how late she was for the meeting and that she had forgotten an important document.
FRIENDLY	sociable, amiable The friendly girl showed the tourists the way to the town centre.
GRATEFUL	thankful Kelly was very grateful for the kindness shown by the stranger.
GUILTY	feeling sorry for doing something wrong Charlie felt guilty about having an affair.
HATEFUL	showing intense dislike

	The two sisters were hateful to each other and always fighting.
HOPEFUL	optimistic Larry was hopeful that the post would bring good news.
HORRIFIED	terrified, appalled The man was horrified to discover that his new wife was already married.
HOSTILE	unfriendly The two neighbours were hostile towards each other because of an argument about loud music.
IMPATIENT	restless, wanting something to happen soon Jane grew increasingly impatient as she waited for her friend who was already 20 minutes late.
IMPLORING	begging, pleading Nicola looked imploring as she tried to persuade her dad to lend her the car.
INCREDULOUS	not believing Simon was incredulous when he heard that he had won the lottery.
INDECISIVE	unsure, hesitant, unable to make your mind up Tammy was so indecisive that she couldn't even decide what to have for lunch.
INDIFFERENT	disinterested, unresponsive, don't care Terry was completely indifferent as to whether they went to the cinema or the pub.
INSISTING	demanding, persisting, maintaining After a work outing, Frank was insisting he paid the bill for everyone.
INSULTING	rude, offensive The football crowd was insulting the referee after he gave a penalty.
INTERESTED	inquiring, curious After seeing Jurassic Park, Hugh grew very interested in dinosaurs.

INTRIGUED	very curious, very interested A mystery phone call intrigued Zoe.
IRRITATED	exasperated, annoyed Frances was irritated by all the junk mail she received.
JEALOUS	envious Tony was jealous of all the taller, better-looking boys in his class.
JOKING	being funny, playful Gary was always joking with his friends.
NERVOUS	apprehensive, tense, worried Just before her job interview, Alice felt very nervous.
OFFENDED	insulted, wounded, having hurt feelings When someone made a joke about her weight, Martha felt very offended.
PANICKED	distraught, feeling of terror or anxiety On waking to find the house on fire, the whole family was panicked.
PENSIVE	thinking about something slightly worrying Susie looked pensive on the way to meeting her boyfriend's parents for the first time.
PERPLEXED	bewildered, puzzled, confused Frank was perplexed by the disappearance of his garden gnomes.
PLAYFUL	full of high spirits and fun Neil was feeling playful at his birthday party.
PREOCCUPIED	absorbed, engrossed in one's own thoughts Worrying about her mother's illness made Debbie preoccupied at work
PUZZLED	perplexed, bewildered, confused After doing the crossword for an hour, June was still puzzled by one clue.
REASSURING	supporting, encouraging, giving someone confidence Andy tried to look reassuring as he told his wife that her new dress did suit her.

REFLECTIVE	contemplative, thoughtful George was in a reflective mood as he thought about what he'd done with his life.
REGRETFUL	sorry Lee was always regretful that he had never travelled when he was younger.
RELAXED	taking it easy, calm, carefree On holiday, Pam felt happy and relaxed.
RELIEVED	freed from worry or anxiety At the restaurant, Ray was relieved to find that he had not forgotten his wallet.
RESENTFUL	bitter, hostile The businessman felt very resentful towards his younger colleague who had been promoted above him.
SARCASTIC	cynical, mocking, scornful The comedian made a sarcastic comment when someone came into the theatre late.
SATISFIED	content, fulfilled Steve felt very satisfied after he had got his new flat just how he wanted it.
SCEPTICAL	doubtful, suspicious, mistrusting Patrick looked sceptical as someone read out his horoscope to him.
SERIOUS	solemn, grave The bank manager looked serious as he refused Nigel an overdraft.
STERN	severe, strict, firm The teacher looked very stern as he told the class off.
SUSPICIOUS	disbelieving, suspecting, doubting After Sam had lost his wallet for the second time at work, he grew suspicious of one of his colleagues.
SYMPATHETIC	kind, compassionate The nurse looked sympathetic as she told the patient the bad

	news.
TENTATIVE	hesitant, uncertain, cautious Andrew felt a bit tentative as he went into the room full of strangers.
TERRIFIED	alarmed, fearful The boy was terrified when he thought he saw a ghost.
THOUGHTFUL	thinking about something Phil looked thoughtful as he sat waiting for the girlfriend he was about to finish with.
THREATENING	menacing, intimidating The large, drunken man was acting in a very threatening way.
UNEASY	unsettled, apprehensive, troubled Karen felt slightly uneasy about accepting a lift from the man she had only met that day.
UPSET	agitated, worried, uneasy The man was very upset when his mother died.
WORRIED	anxious, fretful, troubled When her cat went missing, the girl was very worried.

APPENDIX E: WORKING MEMORY TASK

<u>Instructions:</u>

You will see a sentence on the screen and hear it being said aloud. Your job is to read the sentence out loud along with me, and then as soon as you have finished reading the sentence, decide if the sentence is True or False by checking off either True or False on your sheet of paper that you have in front of you.

After we read some sentences, you will see a screen that says, "What was the last word in each sentence that you read from set #1, 2, 3 etc." When you see this screen, you will be prompted to write down the last word of each sentence that we read from that specific set of words. Don't worry about spelling!

Please put your pencils down as soon as you have finished writing down the words.

Let's give it a try…

<u>Practice Question</u>

A) Jackets have a **zipper**.

TRUE FALSE

B) Shoes go on your **hand**.

TRUE FALSE

What was the last word in each sentence you read? Please put pencils down after writing down the words.

____________________ _______________________

Let's do some more.

<u>Set #1</u>

1) The sun rises in the morning.

TRUE FALSE

2) Trees lose their leaves in spring.

TRUE FALSE

What was the last word of each sentence you read for Set #1?

_____________________ _______________________

<u>Set #2</u>

1) A race car is fast.

TRUE FALSE

2) Peas are vegetables.

TRUE FALSE

What was the last word of each sentence you read for Set #2?

____________________ ______________________

Set #3

1) Dogs have six legs.

TRUE FALSE

2) Giraffes are tall.

TRUE FALSE

What was the last word of each sentence you read for Set #3?

____________________ ______________________

Set #4

1) Cars have four wheels.

TRUE FALSE

2) Cows eat meat.

TRUE FALSE

3) A red traffic light means “STOP”.

TRUE FALSE

What was the last word of each sentence you read for Set #4?

______________________ ________________________ _______________________

Set #5

1) Hens lay eggs.

TRUE FALSE

2) Elephants have purple spots.

TRUE FALSE

3) Stars are in the sky.

TRUE FALSE

What was the last word of each sentence you read for Set #5?

______________________ ________________________ _______________________

Set #6

1) Horses sleep in barns.

TRUE FALSE

2) Boiling water is hot.

TRUE FALSE

3) Strawberries are a fruit.

TRUE FALSE

What was the last word of each sentence you read for Set #6?

____________________ ____________________ ____________________

Set #7

1) We get milk from cows.
   TRUE FALSE
2) Plants need water to grow.
   TRUE FALSE
3) It is warm in Winter.
   TRUE FALSE
4) Carrots are orange.
   TRUE FALSE

What was the last word of each sentence you read for Set #7?

____________________ ____________________ ____________________

____

Set #8

1) Birds have wings.
   TRUE FALSE
2) Whales live in the ocean.
   TRUE FALSE
3) An apple is a fruit.
   TRUE FALSE
4) Fish swim in the sky.
   TRUE FALSE

What was the last word of each sentence you read for Set #8?

____________________ ____________________ ____________________

____

Set #9

1) A soccer ball is round.
   TRUE FALSE
2) We sleep at night.
   TRUE FALSE
3) Bees make honey.
   TRUE FALSE
4) A feather is heavy.
   TRUE FALSE

What was the last word of each sentence you read for Set #9?

____________________ ____________________ ____________________

____

Set #10

1) Birds fly south for the Winter.

TRUE FALSE

2) The earth travels around the sun.

TRUE FALSE

3) Purple is a colour.

TRUE FALSE

4) A car is a vegetable.

TRUE FALSE

5) Tadpoles become frogs.

TRUE FALSE

What was the last word of each sentence you read for Set #10?

______________________ ______________________ ______________________

______________________ ______________________

Set #11

1) Grass is green.

TRUE FALSE

2) Monkeys eat bananas.

TRUE FALSE

3) Pizza is a plant

TRUE FALSE

4) Ice is hot.

TRUE FALSE

5) Basketball is a sport.

TRUE FALSE

What was the last word of each sentence you read for Set #11?

______________________ ______________________ ______________________

______________________ ______________________

Set #12

1) Ants are insects.

TRUE FALSE

2) Lions live on farms.

TRUE FALSE

3) Dogs can bark.

TRUE FALSE

4) Spiders have two legs.

TRUE FALSE

5) A beach has sand.

TRUE FALSE

What was the last word of each sentence you read for Set #12?

_______________ _______________ _______________

_______________ _______________

APPENDIX F: SELF-PERCEPTION SCALES

Items from the following scales were presented in a mixed order to participants within one questionnaire booklet. The below rating scale was provided for each item.

1	2	3	4	5	6
Disagree Strongly	Disagree Moderately	Disagree Slightly	Agree Slightly	Agree Moderately	Agree Strongly

Toronto Empathy Questionnaire
(Spreng, McKinnon, Mar, & Levine, 2009)

1. When someone else is feeling excited, I tend to get excited too
2. Other people's misfortunes do not disturb me a great deal
3. It upsets me to see someone being treated disrespectfully
4. I remain unaffected when someone close to me is happy
5. I enjoy making other people feel better
6. I have tender, concerned feelings for people less fortunate than me
7. When a friend starts to talk about his\her problems, I try to steer the conversation towards something else
8. I can tell when others are sad even when they do not say anything
9. I find that I am "in tune" with other people's moods
10. I do not feel sympathy for people who cause their own serious illnesses
11. I become irritated when someone cries
12. I am not really interested in how other people feel
13. I get a strong urge to help when I see someone who is upset
14. When I see someone being treated unfairly, I do not feel very much pity for them
15. I find it silly for people to cry out of happiness
16. When I see someone being taken advantage of, I feel kind of protective towards him\her

Social Anxiety in University Students Scale
(Bhamani, 2013)

1. I feel comfortable being introduced to new people.
2. I feel comfortable when people watch me doing something.
3. I only have friends of same gender.
4. I feel comfortable with public speaking.
5. I feel comfortable to make friends with people who don’t dress like me.
6. I attend university parties and concerts.
7. I go to my university class fellows’ place for combined studies.
8. I call my university class fellows for combined studies at my place.
9. I feel comfortable to sit next to opposite gender class fellows.

10. I feel comfortable to study from a teacher of opposite gender.
11. My friends think that I am confident.
12. My friends think that I am an extrovert.
13. My friends think that I am easily approachable.
14. My friends feel comfortable to call me at home.
15. My friends think I can easily talk to opposite gender people

Differentiation of Emotions Scale

(Stanovich & West, 2015)

1. My internal states sometimes mystify me.
2. I know when I am angry.
3. I am often confused about my emotional states.
4. Other people seem to know more about what I am feeling that I do.
5. I often think that I am happy but later realize that I wasn’t.
6. I am good at describing my feelings to other people.
7. Sometimes I just don’t know how I feel about people.
8. After watching a movie, I am often confused about how I feel about it.
9. I am aware when I am angry with someone
10. I do not hide my feelings from myself.
11. My emotions are sometimes so mixed up that I don’t know how I am feeling.
12. I am the type of person who is in touch with their emotions.
13. I am good at describing what emotional state I am in.
14. I know when I am happy.

APPENDIX G: COGNITIVE EMOTION REGULATION QUESTIONNAIRE

Please read the instructions below

How do you cope with events?

Everyone gets confronted with negative or unpleasant events now and then and everyone responds to them in his or her own way. By the following questions you are asked to indicate what you generally think, when you experience negative or unpleasant events.

1. I think that I have to accept that this has happened.

1	2	3	4	5	6
Disagree Strongly	Disagree Moderately	Disagree Slightly	Agree Slightly	Agree Moderately	Agree Strongly

2. I often think about how I feel about what I have experienced.

1	2	3	4	5	6
Disagree Strongly	Disagree Moderately	Disagree Slightly	Agree Slightly	Agree Moderately	Agree Strongly

3. I think I can learn something from the situation.

1	2	3	4	5	6
Disagree Strongly	Disagree Moderately	Disagree Slightly	Agree Slightly	Agree Moderately	Agree Strongly

4. I feel that I am the one who is responsible for what has happened.

1	2	3	4	5	6
Disagree Strongly	Disagree Moderately	Disagree Slightly	Agree Slightly	Agree Moderately	Agree Strongly

5. I think that I have to accept the situation.

1	2	3	4	5	6
Disagree Strongly	Disagree Moderately	Disagree Slightly	Agree Slightly	Agree Moderately	Agree Strongly

6. I am preoccupied with what I think and feel about what I have experienced.

1	2	3	4	5	6
Disagree Strongly	Disagree Moderately	Disagree Slightly	Agree Slightly	Agree Moderately	Agree Strongly

7. I think of pleasant things that have nothing to do with it.

1	2	3	4	5	6
Disagree Strongly	Disagree Moderately	Disagree Slightly	Agree Slightly	Agree Moderately	Agree Strongly

8. I think that I can become a stronger person as a result of what has happened.

1	2	3	4	5	6
Disagree Strongly	Disagree Moderately	Disagree Slightly	Agree Slightly	Agree Moderately	Agree Strongly

9. I keep thinking about how terrible it is what I have experienced.

1	2	3	4	5	6
Disagree Strongly	Disagree Moderately	Disagree Slightly	Agree Slightly	Agree Moderately	Agree Strongly

10. I feel that others are responsible for what has happened.

1	2	3	4	5	6
Disagree Strongly	Disagree Moderately	Disagree Slightly	Agree Slightly	Agree Moderately	Agree Strongly

11. I think of something nice instead of what has happened.

1	2	3	4	5	6
Disagree Strongly	Disagree Moderately	Disagree Slightly	Agree Slightly	Agree Moderately	Agree Strongly

12. I think about how to change the situation.

1	2	3	4	5	6
Disagree Strongly	Disagree Moderately	Disagree Slightly	Agree Slightly	Agree Moderately	Agree Strongly

13. I think that it hasn’t been too bad compared to other things.

1	2	3	4	5	6
Disagree Strongly	Disagree Moderately	Disagree Slightly	Agree Slightly	Agree Moderately	Agree Strongly

14. I think that basically the cause must lie within myself.

1	2	3	4	5	6
Disagree Strongly	Disagree Moderately	Disagree Slightly	Agree Slightly	Agree Moderately	Agree Strongly

15. I think about a plan of what I can do best.

1	2	3	4	5	6
Disagree Strongly	Disagree Moderately	Disagree Slightly	Agree Slightly	Agree Moderately	Agree Strongly

16. I tell myself that there are worse things in life.

1	2	3	4	5	6
Disagree Strongly	Disagree Moderately	Disagree Slightly	Agree Slightly	Agree Moderately	Agree Strongly

17. I continually think how horrible the situation has been.

1	2	3	4	5	6
Disagree Strongly	Disagree Moderately	Disagree Slightly	Agree Slightly	Agree Moderately	Agree Strongly

18. I feel that basically the cause lies with others.

1	2	3	4	5	6
Disagree Strongly	Disagree Moderately	Disagree Slightly	Agree Slightly	Agree Moderately	Agree Strongly