



MASTERARBEIT | MASTER'S THESIS

Titel | Title

Comparing Miscalibration and Confidence Judgment: Exploring
the Impact on Decision-Making

verfasst von | submitted by

Benjamin Kopic

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of
Master of Science (MSc)

Wien | Vienna, 2026

Studienkennzahl lt. Studienblatt | Degree
programme code as it appears on the
student record sheet:

UA 066 915

Studienrichtung lt. Studienblatt | Degree
programme as it appears on the student
record sheet:

Masterstudium Betriebswirtschaft

Betreut von | Supervisor:

Assoz. Prof. Steffen Keck PhD

Abstract

Confidence calibration plays a central role in decision-making across healthcare, finance, and education. It describes how well a person's subjective certainty matches their actual accuracy and when this alignment breaks down, overconfidence emerges. This study examines how different types of feedback influence two distinct dimensions of overconfidence, namely overestimation and overplacement. Using a between-subjects design with 84 participants randomly assigned to three groups, the study compared outcome-based feedback, comparative feedback, and a no-feedback control condition across two rounds of a general knowledge quiz. The results demonstrate that outcome-based feedback significantly reduced overestimation within the feedback group. Furthermore, it produced a meaningful but statistically borderline reduction in overplacement, even though it was not specifically designed to target that dimension. Comparative feedback yielded no significant changes on either measure, while the control group showed a slight increase in overplacement over time. These findings suggest that the specificity of feedback matters more than social comparison when it comes to recalibrating confidence. Outcome-based feedback appears to trigger a broader reassessment of self-perception that extends beyond its immediate target. Implications for feedback design in educational, organizational, and healthcare contexts are discussed.

Kurzfassung

Selbstüberschätzung zählt zu den am häufigsten dokumentierten Verzerrungen im menschlichen Urteil. Sie beschreibt die Tendenz, sich bei Entscheidungen sicherer zu fühlen als es die tatsächliche Leistung rechtfertigt, und beeinflusst Ergebnisse in Medizin, Finanzen und Bildung. Diese Studie untersucht, ob verschiedene Arten von Feedback diese

Selbstüberschätzung reduzieren können, mit besonderem Fokus auf zwei ihrer Dimensionen, nämlich Überschätzung (Overestimation) und Überplatzierung (Overplacement). Eine empirische Untersuchung wurde mittels einer Online-Umfrage mit 84 Teilnehmenden durchgeführt, die zufällig drei Gruppen zugewiesen wurden. Die Teilnehmenden absolvierten einen Allgemeinwissenstest in zwei Runden, wobei ihre Selbsteinschätzungen und tatsächlichen Ergebnisse erfasst wurden. Die Studie verglich drei Bedingungen, nämlich ergebnisbasiertes Feedback, vergleichendes Feedback und eine Kontrollgruppe ohne Feedback, um deren Einfluss auf die Kalibrierung des Selbstvertrauens zu bewerten. Die Ergebnisse zeigen, dass ergebnisbasiertes Feedback die Überschätzung innerhalb der Feedbackgruppe signifikant reduzierte und zudem eine inhaltlich bedeutsame, statistisch jedoch grenzwertige Verbesserung der Überplatzierung bewirkte. Vergleichendes Feedback zeigte auf keiner der beiden Dimensionen signifikante Veränderungen. Die Kontrollgruppe wies über die Zeit eine leichte Zunahme der Überplatzierung auf. Die Befunde deuten darauf hin, dass die Spezifität von Feedback bei der Rekalibrierung wichtiger ist als der soziale Vergleich. Die Studie liefert nützliche Einblicke für die Feedbackgestaltung in Bildungs-, Organisations- und Gesundheitskontexten und unterstreicht den Bedarf an weiterer Forschung zu den psychologischen Mechanismen der Selbstüberschätzung.

Table of Contents

Table of Contents	4
1. Introduction.....	6
2. Literature Overview	11
2.1 Historical Overview	11
2.2 Theoretical Foundations of Confidence and Calibration	17
2.3 Feedback and Confidence Calibration	24
2.4 Neuroscience of Confidence and Miscalibration	29
2.5 Practical Implications of Overconfidence and Feedback	32
2.6 Summary and Conceptual Framework.....	36
3. Research Question and Hypotheses	39
3.1 Research Gap	39
3.2 Research Question	40
3.3 Hypothesis Development.....	41
3.4 Measurement of Variables.....	41
4. Experiment Design & Methodology	42
4.1 Survey Structure.....	43
4.2 Experimental Groups	44
4.3 Variable Description.....	45
4.4 Participants and Recruitment	46
5. Results.....	47
5.1 Descriptive Statistics	47
5.2 Assumption Checks	49
5.2.1 Normality	49
5.2.2 Homogeneity of Variance	50
5.3 Between-Group Comparisons.....	50
5.3.1 Hypothesis 1 Overestimation	51
5.3.2 Hypothesis 2 Overplacement	51
5.3.3 Hypothesis 3 Overall Calibration.....	52
5.4 Within-Group Paired Comparisons	52
5.4.1 Group A Outcome-Based Feedback.....	53
5.4.2 Group B Comparative Feedback.....	54
5.4.3 Group C Control.....	54

5.5 Summary of Findings.....	55
6. Discussion	56
6.1 Outcome-Based Feedback and Overestimation	56
6.2 Comparative Feedback and Overplacement.....	57
6.3 Group A and Overplacement	59
6.4 Group A and Quiz Performance	60
6.5 Comparative Feedback Performance.....	60
6.6 Implications.....	61
6.7 Limitations.....	62
7.1 Directions for Future Research	64
7.2 Final Remarks	65
References	66

1. Introduction

Decision-making is part of everyday life and influences many areas, including healthcare, finance, and education. People make choices constantly, and some of these choices can have serious consequences. In healthcare, for example, doctors often need to make decisions quickly and with limited information. These decisions can affect individual patients as well as wider health systems (Masic, 2022). This became especially clear during the COVID-19 pandemic, when knowledge was limited and many decisions, particularly those related to vaccines, had to be made very quickly.

A similar situation exists in finance, where investors must decide how to allocate their money in markets that are uncertain and often unstable. These decisions are influenced not only by available information but also by behavioral patterns that are well documented in behavioral economics (Carminati, 2020). These examples help explain why understanding both decision-making and confidence is important.

Confidence plays an important role in how decisions are made. It affects how people interpret situations, justify their choices, and act on them. It also shapes how others perceive their expertise. Patients, for instance, often trust doctors who appear confident, and investors who seem confident are frequently viewed as more capable, even when their predictions are not more accurate than those of others. In organizations, confident leaders can motivate their teams. Still at the same time, high confidence can encourage riskier decisions that may lead to negative outcomes. For this reason, confidence is a complex and sometimes problematic part of human judgment.

Many decisions are based on intuition, because people often rely on fast, experience-based judgments to handle complex situations. Intuition can be useful, but it is not always

reliable. In particular, the confidence that supports quick decisions is sometimes misplaced and people can feel certain about their choices even when those choices later turn out to be wrong, and they may repeat the same mistakes. The gap between how confident someone feels and how accurate they actually are is known as miscalibration. This gap can be especially harmful when decisions involve high stakes.

The relationship between confidence and accuracy has been studied for many decades. In the 1950s, Paul Meehl showed that statistical models often outperform human experts, suggesting that people tend to be more confident than they should be (Meehl, 1954). His work raised early questions about how reliable intuition is compared to data-driven approaches. In the 1970s, Tversky and Kahneman introduced the concept of heuristics, which are mental shortcuts that help people make faster decisions but can also lead to systematic biases (Tversky & Kahneman, 1974). Their research showed that individuals often overestimate their own knowledge and abilities.

Since then, research in psychology, behavioral economics, and neuroscience has expanded our understanding of how people judge their own decisions. Confidence biases influence financial markets, risk-taking behavior, and organizational decision-making. Overconfident investors, for example, may trade too often because they believe they can outperform the market, while underconfident individuals, on the other hand, may hesitate and miss opportunities. Both patterns highlight the importance of balancing confidence and accuracy (Barber & Odean, 2001).

This need for balance is also seen in applied fields such as medicine, education, and aviation. In healthcare, diagnostic errors linked to overconfidence have increased interest in how professionals use and respond to feedback. In education, asking students to reflect on their level of certainty has been shown to support learning (Dunlosky & Metcalfe, 2009). In aviation,

pilots receive constant performance feedback, which helps them maintain better calibration (Endsley, 1995). Across these fields, the challenge is not only to make accurate decisions but also to judge how accurate those decisions are likely to be. Feedback plays an important role in addressing this challenge.

Feedback helps people adjust their confidence over time. When individuals receive clear information about whether they were right or wrong, they can update their expectations in future decisions. Over time, this can reduce the gap between perceived and actual performance. When feedback is missing, delayed, or unclear, miscalibration often persists. This helps explain why experts who receive regular feedback are usually better calibrated than beginners and even so, experts can still make mistakes in environments where feedback is limited or misleading (Koriat, 2012).

Accepting feedback is not always easy because emotional and motivational factors strongly influence how people interpret it. Many individuals accept feedback that supports their self-view but resist feedback that challenges it (Dunning, 2011). In professional settings where confidence is closely linked to competence, admitting mistakes can feel uncomfortable and defensive reactions can limit learning and slow improvement (Baumeister, 1999). To better understand how feedback interacts with these reactions, it is useful to look at different forms of overconfidence.

Moore and Healy (2008) argued that overconfidence is not a single concept but consists of three related forms: overestimation, overplacement, and overprecision. Overestimation refers to believing that one performed better than they actually did. Overplacement describes the belief that one is better than others, such as assuming one is a safer driver than the average person. Overprecision refers to being overly certain about one's beliefs, even when the evidence is weak (Moore & Healy, 2008)

Each form of overconfidence has different causes and consequences. Overestimation can lead to unrealistic expectations and poor preparation while overplacement may create social or professional tensions. At the other end overprecision can make individuals resistant to corrective information. These differences provide a useful framework for examining how feedback can improve confidence calibration (Klayman et al., 1999).

Recent research suggests that the way feedback is delivered also matters. Simple right or wrong feedback is often less effective than more detailed information that shows how far a judgment was from the correct answer. Emotional and motivational responses further influence whether feedback is used constructively and in some cases people learn from it, while others ignore or reject it. This suggests that feedback alone is not enough to eliminate overconfidence and its effects depend on individual differences in motivation, learning style, and self-awareness (Lichtenstein & Fischhoff, 1977).

Beyond feedback, confidence is shaped by personal experience and cultural context. Confidence judgments are influenced not only by cognitive processes but also by cultural norms, especially in ones that value assertiveness, overplacement may be more common. In cultures that emphasize modesty, underconfidence may be more frequent (Heine et al., 1999). Work environments also differ. Pilots, surgeons, and financial analysts often receive frequent and high-stakes feedback, whereas teachers or policymakers may receive slower or less direct information. These differences show that confidence is shaped by both individual thinking and the environment in which decisions are made.

The central challenge is to align confidence with actual accuracy. Overconfidence can lead to serious consequences, such as misdiagnoses or financial losses. Underconfidence can lead to hesitation and missed opportunities. Both outcomes reflect failures of calibration.

Improving calibration requires reliable and timely feedback that helps individuals adjust how confident they feel about their decisions (Highhouse, 2008).

Although overconfidence has been studied for many years, we still know less about how different types of feedback influence its different forms. Many studies examine feedback in a general way, and several focus mainly on overprecision. In contrast, overestimation and overplacement are often not clearly separated. Outcome based and comparative feedback are also rarely tested within the same experiment. Because of this, it remains unclear whether these feedback types influence confidence calibration in similar ways or whether they affect different psychological mechanisms, especially those linked to motivation and social comparison.

This gap makes a more structured comparison necessary. The present thesis therefore investigates how different feedback types influence confidence calibration, with a specific focus on overestimation and overplacement. By directly comparing outcome based feedback and comparative feedback within one unified experimental design, the study aims to clarify whether each feedback type reduces a different dimension of overconfidence. Rather than treating calibration as a single overall measure, the analysis distinguishes between absolute misjudgment of performance and relative misjudgment compared to others.

Based on prior research, it is expected that individuals who receive feedback will show better calibration than those who receive no feedback. However, the effects are not expected to be identical across dimensions. Outcome based feedback should primarily reduce overestimation, because it directly reveals the gap between perceived and actual performance. Comparative feedback, on the other hand, should be more effective in reducing overplacement, since it provides information about relative standing within a group. In summary, this thesis examines how individuals interpret and respond to different forms of feedback and how these responses shape confidence and decision accuracy over time. By separating overestimation

from overplacement and comparing feedback types within the same framework, the study aims to provide a clearer understanding of confidence calibration and offer practical implications for improving decision quality in areas such as healthcare, finance, and education.

2. Literature Overview

2.1 Historical Overview

Research on confidence and decision accuracy has developed steadily over the past century. What began as an interest in judgment and intuition has grown into an interdisciplinary field spanning psychology, economics, and neuroscience. Despite this development, the central question has remained largely the same: how well do people understand the limits of their own knowledge?

Early researchers tried to explain why individuals often feel highly confident about decisions that later turn out to be wrong. Over time, this work led to the concepts of miscalibration and overconfidence, which continue to shape research in this field (Lichtenstein & Fischhoff, 1980).

This question remains important because people regularly face situations in which they must judge both what they know and how certain they feel about it. Whether in medicine, finance, or everyday life, decisions are often made under uncertainty. Understanding confidence helps explain why individuals sometimes rely too strongly on their own judgment, while in other cases they underestimate their abilities. These patterns can strongly influence outcomes across different domains.

Systematic research on judgment errors began in the 1950s. One of the most influential early contributions was Paul Meehl's book *Clinical versus Statistical Prediction* (1954). Meehl showed that simple statistical models often outperformed human experts when predicting outcomes such as academic success or clinical diagnoses. He observed that experts were typically very confident in their assessments but frequently underestimated the limits of their intuition. His findings revealed a clear gap between expected and actual accuracy, challenging the long-standing belief that expert intuition was naturally superior to statistical reasoning.

Meehl's work also had implications beyond academic psychology. It encouraged the use of structured assessment tools and evidence-based methods in fields such as medicine, education, and personnel selection. Over time, many institutions shifted away from relying mainly on intuition and toward more systematic and data driven decision-making approaches (Dawes, Faust & Meehl, 1989). In this way, Meehl's ideas contributed to a broader discussion about how decisions should be made in complex and uncertain situations.

Following Meehl's work, researchers increasingly focused on the psychological mechanisms that cause people to misjudge their own accuracy. During the 1960s and 1970s, research expanded to include the mental processes that shape confidence and error. Overconfidence was no longer seen as a rare mistake but as a common feature of human thinking. People were found to rely on mental shortcuts that simplify complex problems. While these shortcuts allow for faster decisions, they also lead to systematic biases (Adams & Adams, 1961).

This line of research led to the influential work of Amos Tversky and Daniel Kahneman. Their studies on heuristics and biases fundamentally changed how human decision-making was understood. In their well-known 1974 paper *Judgment under Uncertainty*, they explained how individuals use intuitive rules of thumb to judge probabilities and predict outcomes. These

shortcuts make decisions faster, but they also produce consistent errors. For example, the availability heuristic leads people to judge events as more likely when examples come easily to mind, while the representativeness heuristic causes judgments to be based on similarity rather than statistical reasoning. Together, these processes help explain why individuals can feel highly confident even when their beliefs are inaccurate (Tversky & Kahneman, 1974). This is important for the present study, as it suggests that feedback must correct or reshape these intuitive judgments if confidence is to improve.

This bias also influenced applied areas such as forecasting, policy design, and risk communication. Governments and organizations began using psychological insights to better understand how people respond to uncertainty, including in health communication and financial decision-making. This period marked an important shift in applied psychology, as research on judgment biases started to inform the design of decision systems rather than only describing human error.

Around the same time, Herbert Simon introduced the concept of bounded rationality (1957), which provided a useful framework for explaining these findings. Simon argued that human decision-making is constrained by limited information, limited time, and limited cognitive resources. Rather than identifying optimal solutions, people usually choose options that are good enough given their circumstances. Within this framework, confidence judgments reflect how individuals evaluate their own decisions under uncertainty. Overconfidence can therefore be understood not only as an error but also as a natural outcome of attempting to act efficiently in complex environments.

Bounded rationality also connects closely to later research on cognitive load and decision fatigue. Studies in these areas show that mental effort is limited and can be exhausted relatively quickly (Kahneman, 2011). When individuals face too many decisions or too much

information, their ability to judge confidence accurately declines. They may rely more heavily on intuition, make quicker choices, or behave inconsistently. This supports the idea that confidence is shaped by both cognitive capacity and situational demands. To summarize, while Simon (1957) established the structural limits of human rationality, it was the subsequent work of Tversky and Kahneman (1974) revealed the specific heuristics that explain why these limits are crossed.

As research progressed, psychologists began examining how judgment biases appear outside controlled laboratory settings. During the 1980s, behavioral decision theory emerged as a field that combined psychological insights with economic models (Slovic, Fischhoff & Lichtenstein, 1982). Researchers such as Richard Thaler, Daniel Kahneman, and Amos Tversky challenged the classical assumption that individuals always behave as rational decision-makers who maximize utility. Instead, they showed that emotions, heuristics, and social influences often lead to predictable deviations from rational behavior. Overconfidence became one of the most widely studied biases because it helped explain real-world outcomes such as stock market bubbles, entrepreneurial failures, and overly optimistic business planning. This marked an important shift and the focus moved from asking whether people are biased to understanding how such biases operate in real decision environments.

These ideas later influenced applied fields such as behavioral finance, organizational psychology, and public policy. Thaler's work on nudge theory, for example, demonstrated that small changes in decision environments can help individuals make better and more calibrated choices. This showed that understanding confidence and judgment biases has clear practical value for designing policies and systems that support better everyday decision-making.

Evidence from different domains illustrates the effects of overconfidence. In financial decision-making, overconfident investors tend to overestimate their predictive abilities, trade

excessively, and underestimate risk, which often leads to lower returns (Odean, 1998). For example, an investor who attributes gains to skill but losses to bad luck may never fully adjust their confidence, even after repeated underperformance. In entrepreneurship, founders frequently overestimate their chances of success, contributing to high failure rates among start-ups (Camerer & Lovallo, 1999). In clinical contexts, physicians may overestimate diagnostic accuracy, sometimes resulting in misdiagnosis or delayed treatment (Berner & Graber, 2008). Across these examples, a common pattern appears: confidence remains inflated when feedback is unclear or delayed. While moderate confidence can be motivating, weak feedback allows inaccurate self-assessments to persist.

During the same period, researchers also began to examine how confidence could be improved rather than focusing only on why it is biased. This led to the concept of calibration, which describes how closely confidence matches actual performance. Studies in the 1980s showed that people are often poorly calibrated. For example, when individuals reported being “80 percent sure,” their actual accuracy was typically closer to 60 or 70 percent (Arkes, Dawes & Christensen, 1986).

Researchers then examined whether feedback could reduce this gap and if it allows individuals to adjust future judgments. Stone (1994) showed that immediate and consistent feedback reduced overestimation in simple tasks. It was found that repeated feedback over time helped participants align their confidence more closely with their actual performance. These findings highlighted feedback as an important learning signal linking beliefs to objective outcomes. However, most of these early studies were based on relatively simple estimation tasks, which leaves open how feedback functions in more complex decision settings.

The concept of calibration soon influenced education and professional training. In schools and universities, self-testing and formative assessment were introduced to help learners

compare perceived and actual knowledge. In high-stakes professions such as aviation, medicine, and surgery, simulation-based training combined with feedback loops improved both performance and confidence calibration. These applications demonstrated how psychological research could inform practical training methods.

By the late twentieth century, researchers also recognized that overconfidence is not a single phenomenon. Moore and Healy (2008) proposed three distinct forms: overestimation, overplacement, and overprecision. This framework helped explain inconsistencies in earlier findings. Later studies suggested that feedback reduces overestimation most strongly, has moderate effects on overplacement, and has the weakest impact on overprecision, which appears to be the most resistant to change. Since these findings often come from different experimental designs, it remains unclear whether the differences reflect real psychological mechanisms or methodological variation across studies which were done. This framework became influential because it allowed researchers to study confidence in a more structured way. By separating cognitive, social, and perceptual components, it became easier to examine how different contexts and motivations influence calibration.

At the same time, interest in the biological basis of confidence increased. Advances in neuroscience made it possible to examine how the brain forms and updates confidence judgments. Research by Fleming and Dolan (2012) showed that regions of the prefrontal cortex play a key role in generating confidence. When expectations do not match outcomes, error-detection regions such as the anterior cingulate cortex become active and help adjust future judgments. These findings align with computational models suggesting that confidence functions like a probability estimate based on uncertain evidence (Kepecs & Mainen, 2012).

In recent years, confidence research has expanded beyond laboratory experiments into applied settings. In medicine, diagnostic overconfidence contributes to avoidable errors, but

structured audits and peer feedback systems can reduce this bias (Balla & Heneghan, 2011). In education, self-testing and reflective practices help students identify gaps between confidence and knowledge, improving both learning and self-awareness. Organizations also use feedback and performance evaluations to help employees better understand their strengths and limitations (Zimmerman & Schunk, 2011).

Digital tools have also become part of this development since simulation environments and adaptive learning systems now provide frequent feedback that supports calibration in real time. These tools rely on principles from cognitive research and show how earlier psychological findings continue to shape modern training and decision-support methods.

Overall, the historical development of confidence research reflects a sustained effort to understand how people evaluate their own knowledge and decisions. While individuals often overestimate their accuracy, research consistently shows that clear and reliable feedback can improve calibration. This historical overview provides the foundation for the theoretical models discussed in the following sections.

2.2 Theoretical Foundations of Confidence and Calibration

Building on the ideas of Tversky and Kahneman, modern research focuses on the specific mistakes in thinking that cause "miscalibration." The most common one is overprecision. This is when someone is too sure that their answers or beliefs are correct.

However, to really understand how people judge their own confidence, we need to look at three main areas: psychology, economics, and neuroscience. Research in these fields shows that how "sure" we feel is often different from how "right" we actually are. Also, people don't

all learn from feedback in the same way (Fleming & Dolan, 2012). Instead of being a perfect reflection of the truth, confidence is something our brain creates based on our past experiences.

Together, these ideas explain why confidence sometimes improves with experience, but not always for everyone. Differences in how we process information and monitor our own thoughts can lead to "persistent miscalibration," even for experts (Kahneman, 2011). Over the last fifty years, research has focused on four main areas: signal detection theory, dual-process theory, metacognitive monitoring, and feedback-based learning. Together, these help us see confidence as a process that is constantly being updated by our environment.

Signal detection theory, developed in the mid twentieth century to study perceptual decision making under uncertainty, remains one of the most widely used models for analyzing confidence. At its core, the theory assumes that decisions arise from comparing internal evidence against a decision threshold. When the internal signal exceeds this threshold, the observer decides that a stimulus is present. When it does not, the observer decides it is absent. Confidence is coming from the same evidence distribution that supports the decision itself. Stronger evidence leads to higher subjective certainty, while weak evidence results in uncertainty. This framework established a crucial link between accuracy and confidence, showing that both depend on the strength and reliability of available evidence (Macmillan & Creelman, 2005).

In applied settings, the relationship between evidence strength and confidence is rarely proportional. Decision making involves perceptual noise, motivational influences, and uneven incentives that are shifting thresholds away from their optimal values. For example, a physician who repeatedly makes correct diagnoses based on a specific symptom pattern may gradually become more confident when encountering that pattern again. Over time, this can increase the risk of overconfidence in similar future cases. Signal detection theory helps explain

how such shifts occur and it provides a framework for examining how well individuals distinguish between correct and incorrect decisions. Measures such as the ROC curve or the Brier score are used to assess whether confidence levels correspond to actual accuracy (Macmillan & Creelman, 2005). In applied contexts, these measures are used to evaluate predictive performance in both human judgment and artificial intelligence systems. Findings suggest that when decisions are made under uncertainty, miscalibration is common in both. This indicates that discrepancies between confidence and accuracy are not only individual shortcomings but a broader feature of systems operating under uncertain conditions.

Signal detection theory has also been extended beyond basic perception to higher order decision processes, including memory retrieval, moral reasoning, and diagnostic judgment. These applications model confidence as a form of probabilistic inference in which individuals estimate not only what is likely to be true but also how certain they are about that conclusion. This interpretation connects psychological research with computational models in neuroscience and suggests that confidence reflects an internal signal of evidence strength that is updated through experience and feedback (Kepecs & Mainen, 2012).

While signal detection theory captures the statistical structure of decision evidence, it does not explain how confidence is subjectively experienced or cognitively regulated. This gap is addressed by dual process theory, most prominently described in Daniel Kahneman's model of System 1 and System 2 thinking. According to this framework, cognition operates through two interacting modes. One is intuitive and automatic, relying on pattern recognition and affective cues. The other is slower and reflective, involving analytical reasoning and error checking. System 1 produces rapid feelings of knowing and generates confidence quickly, often before conscious reasoning begins. System 2 monitors these intuitive judgments and revises them when necessary. Because engaging System 2 requires effort and motivation,

individuals often rely on intuitive confidence even in situations where deeper reflection would improve accuracy.

Empirical research supports this distinction. Confidence judgments often emerge before deliberate reasoning. Based on the studies, Glöckner and Betsch (2008) demonstrated that intuitive confidence signals appear milliseconds before analytical evaluation. Following this Thompson and colleagues (2011) showed that people experience a feeling of rightness that predicts confidence even when answers are incorrect. These findings help explain why feedback plays such a central role and without explicit outcome information, intuitive overconfidence remains largely unchallenged. In addition to this it is very important to understand that feedback acts as an external signal that activates analytical processing by indicating that an error has occurred. Over time, repeated feedback allows intuitive mechanisms to be reshaped so that rapid judgments become more closely aligned with analytical accuracy.

Dual process models also help explain how experience influences calibration. Novices, who lack an extensive history of feedback, tend to rely more on intuitive processing and are therefore more prone to miscalibration. Experts, in contrast, develop more reliable intuitions over time because repeated feedback gradually shapes and corrects their automatic responses. This suggests that confidence is not merely a stable personality trait, but a capacity that develops through learning and experience (Kahneman & Klein, 2009). In this sense, expertise involves not only acquiring knowledge, but also improving the ability to judge own accuracy. Because such improvement depends on the capacity to evaluate own judgments, the following section examines the processes underlying metacognitive monitoring.

Metacognitive monitoring looks at how people judge their own thinking. It focuses on how individuals reflect on their decisions and evaluate whether they were correct. The concept

was first introduced by Flavell in 1979 and later developed by Nelson and Narens in 1990, who described confidence as a second level judgment. In other words, confidence is not the decision itself, but an evaluation of how accurate that decision was.

The framework separates two key processes. Monitoring means assessing how well we performed. Control means adjusting our behavior based on that assessment. In real world decision making, people compare what they expected to happen with what actually happened, and they use this comparison to guide future choices. Sometimes this happens consciously and sometimes it does not.

Problems arise when monitoring is inaccurate. If people misjudge their own performance, their control processes are built on distorted evaluations, which can lead to repeated mistakes over time. They may believe they performed well even when they did not. As a result, they fail to notice the gap between perceived and actual performance, and overconfidence can persist. To examine how accurately individuals evaluate their own performance, researchers often use calibration curves that plot confidence against actual accuracy. Perfect calibration appears as a diagonal line, where confidence corresponds to performance. Empirical studies show, however, that results frequently lie above this line, indicating a tendency toward overconfidence (Lichtenstein et al., 1982).

Brunswik's Lens Model describes judgment as a process in which individuals rely on environmental cues, whose validity may be overestimated or underestimated. People sometimes place too much weight on cues that are only weakly related to actual outcomes while overlooking more relevant information. In this context, feedback serves as an important corrective mechanism by showing which cues were predictive and which were not (Klayman et al., 1999). Over time, this process can support more improved calibration.

A well known extension of this perspective is the Dunning Kruger effect (Kruger & Dunning, 1999). Individuals with lower competence often overestimate their performance because they lack the insight needed to recognize their own errors. By contrast, highly skilled individuals may underestimate their relative standing, assuming that tasks they find manageable are similarly manageable for others. These findings suggest that accurate self evaluation depends not only on knowledge, but also on awareness of own limitations.

Finally, emotional states can influence confidence judgments. Research suggests that positive emotions such as pride may increase confidence, whereas anxiety and uncertainty can reduce it (Koriat, 2012). The way feedback is experienced therefore plays a role in learning. Feedback perceived as constructive can encourage adjustment, while feedback experienced as threatening may lead to defensiveness and maintain overconfidence (Baumeister, 1999).

Neuroscientific research complements psychological models by identifying the neural mechanisms through which feedback supports learning. Evidence from functional imaging and electrophysiological studies shows that feedback triggers some neural responses associated with reward processing and error detection. The anterior cingulate cortex plays a key role in detecting discrepancies between expected and actual outcomes, while dopaminergic prediction error signals in the basal ganglia contribute to updating confidence in response to these discrepancies (Schultz, 2016). The prefrontal cortex, particularly medial and dorsolateral regions, supports metacognitive monitoring by integrating internal beliefs with external feedback (Fleming & Dolan, 2012).

These perspectives provide a combined account of confidence formation. Signal detection theory describes the statistical relationship between evidence and decision thresholds. Dual process theory explains how intuitive and analytical systems contribute to the generation and regulation of confidence. Metacognitive theory addresses how individuals monitor and

adjust their judgments, while feedback based learning highlights the motivational and neural mechanisms that enable recalibration. From this perspective, confidence is not a fixed conviction, but a dynamic judgment that is continuously shaped through interaction between cognition and experience (Moore & Healy, 2008). Bringing these approaches together helps explain why different types of overconfidence react differently to feedback. Not all forms of overconfidence change in the same way.

Overestimation, for example, often decreases when people receive clear outcome feedback that shows them they were wrong. Seeing a direct mistake can correct inflated self views. Overplacement is different. Because it is based on social comparison, it requires feedback about how one performed relative to others, not just whether the answer was correct.

Overprecision is even more difficult to change. It refers to being too certain about the accuracy of one's judgments. Simple right or wrong feedback usually does not provide information about how uncertain a situation really was. As a result, overprecision tends to be more resistant to correction. Reducing it may require explicit exposure to probabilities or training that focuses directly on uncertainty, rather than relying only on binary performance feedback.

Overall, these theoretical perspectives suggest that feedback serves as a central link between cognitive evaluation and behavioral adjustment. When feedback is timely, specific, and perceived as constructive, it can encourage reflection and support the revision of confidence judgments. In contrast, when feedback is delayed, ambiguous, or experienced as threatening, learning becomes more difficult and biases are more likely to persist. These foundations indicate that confidence is not a fixed belief, but a dynamic judgment shaped by experience and emotion. Feedback therefore does not automatically improve calibration, but

its effectiveness depends on timing, content, and emotions. The next section examines how these characteristics influence whether individuals successfully adjust their confidence.

2.3 Feedback and Confidence Calibration

Feedback is a powerful and complex mechanism for improving decision making. It functions as both a cognitive and social process, linking perceived performance to actual outcomes. Although feedback often facilitates learning and can reduce bias, it does not automatically eliminate overconfidence. Research in psychology and organizational behavior suggests that the effectiveness of feedback depends on factors such as timing, clarity, and the emotional responses it evokes. Understanding these factors is essential for explaining why some individuals adjust their confidence following feedback, whereas others remain persistently miscalibrated (Koriat, 2012).

Feedback is not merely a technical transmission of information, but also a social and emotional experience. It requires individuals to reconcile internal self evaluations with external performance indicators, a process that can be psychologically challenging. The interaction between cognition and emotion helps explain why feedback can be both informative and difficult to accept. For organizations and educators, the central challenge is to create environments that promote accountability while limiting defensiveness, thereby supporting ongoing learning and development (Baumeister et al., 2001).

At a basic level, feedback functions as an error correction mechanism. It allows individuals to compare their expectations with actual outcomes and adjust their behavior accordingly. This process is similar to principles from reinforcement learning, which explain how differences between expected and observed results drive adaptation. Positive feedback

strengthens confidence in existing strategies, whereas negative feedback signals that adjustments may be necessary. Over repeated cycles, feedback therefore serves not only as information about performance, but also as a learning signal that contributes to the development of metacognitive awareness (Schultz, 2016).

Early research suggests that feedback is most effective when it is timely and clear. Stone (1994) showed that participants who received immediate feedback during probability estimation tasks achieved larger reductions in overestimation than those who received delayed or aggregated feedback. Similarly, Arkes et al. (1987) found that consistent feedback in forecasting tasks helped participants bring their confidence ratings closer to their actual success rates. These findings indicate that the impact of feedback depends on how clearly individuals can connect their actions with resulting outcomes. When this connection is direct and immediate, learning tends to improve, but when it is delayed or unclear, the corrective effect of feedback is reduced.

The impact of feedback on confidence calibration depends strongly on its timing and format. Immediate feedback delivered shortly after a decision typically has a stronger effect because individuals can still recall the reasoning underlying their choice. Klayman et al. (1999) found that trial by trial feedback enabled participants to adjust their confidence dynamically and reduce overconfidence after only a few repetitions. In contrast, delayed feedback may weaken the perceived connection between an action and its outcome, making it more difficult to identify the source of error. However, delay does not necessarily undermine learning. When individuals are required to reconstruct their reasoning before reviewing the outcome, delayed feedback can support deeper reflection. In educational settings, this process promotes metacognitive engagement, as learners must examine how their initial reasoning diverged from correct principles.

Another common format is summary feedback, which aggregates performance across multiple decisions. This approach is useful for identifying general patterns, such as a consistent tendency toward overconfidence, but it provides limited guidance for moment to moment calibration. Larrick (2004) observed that while summary feedback helped individuals adjust their overall confidence levels, they often continued to commit specific judgment errors. Summary feedback therefore supports broader self awareness but offers less precision in correcting local mistakes.

Comparative or social feedback communicates how an individual's performance compares with that of peers. Research by Tenney et al. (2008) suggests that such feedback can reduce overplacement, defined as the belief that one performs better than others, although the effects are often modest. Because social comparison may threaten self esteem, individuals sometimes interpret comparative feedback defensively in order to maintain a positive self image. In some cases, learning that one has performed below average may even reinforce self protective confidence. This is present especially in the cases, when feedback is public.

Beyond structure and timing, feedback processing involves both cognitive and emotional components. Cognitive processing concerns the identification of discrepancies between expected and observed outcomes, whereas emotional processing relates to how individuals respond to these discrepancies. Neuroscientific research indicates that feedback activates brain regions associated with monitoring and self regulation, particularly the anterior cingulate cortex and prefrontal areas (Fleming & Dolan, 2012). These systems generate internal signals that inform subsequent behavior. Whether these signals translate into improved calibration, however, depends largely on the emotional responses they elicit.

Emotion plays an important role in determining whether feedback promotes learning or defensiveness. Positive feedback typically reinforces confidence and motivation, whereas

negative feedback may activate threat related responses that reduce reflective processing. Individuals often respond more strongly to negative than to positive information, which can support error correction but may also encourage avoidance.

When feedback threatens self esteem, individuals may dismiss or reinterpret it in order to preserve a positive self image, thereby sustaining overconfidence. In contrast, when feedback is delivered constructively and framed as an opportunity for development, it is more likely to encourage metacognitive engagement and support calibration (Dunning, 2011).

Feedback Intervention Theory offers an important framework for understanding these dynamics (Kluger & DeNisi, 1996). At its core, the theory proposes that feedback influences performance by directing attention toward different levels of evaluation, including the task and the self. When feedback focuses on the task and provides clear guidance, learning tends to improve because attention remains on problem solving. When it shifts attention toward the self and implies incompetence, defensiveness may emerge. In such cases, engagement with the task declines. Explaining why an answer is incorrect supports recalibration, whereas labeling an individual as incapable undermines it. Effective feedback therefore requires a careful balance between informational clarity and psychological safety.

These principles also help explain why different forms of overconfidence respond differently to feedback. Overestimation, defined as the gap between perceived and actual performance, tends to respond well to factual outcome information, particularly when feedback is consistent and specific. Fischer and Budescu (2005) showed that numerical accuracy feedback reduced this discrepancy over time. Overplacement, the belief that one performs better than others, is more resistant because it involves social comparison rather than simple error correction. It requires explicit and non threatening information about relative standing in order to produce meaningful adjustment (Pulford & Colman, 1997).

Overprecision, the tendency to express excessive certainty, is the most difficult to change. Soll and Klayman (2004) found that individuals continued to report narrow confidence intervals even after repeated errors. Most feedback signals only correctness and it rarely communicates the degree of uncertainty involved. As a result, individuals seldom learn how wide their confidence ranges should be.

The context in which feedback is delivered further shapes its impact. In education, formative feedback supports self regulated learning and gradually improves calibration. In healthcare, audit systems and simulation training provide structured outcome information that can improve diagnostic accuracy, although hierarchical norms and fear of blame may limit openness to correction. In financial decision making, feedback is often indirect and difficult to interpret. Gains are frequently attributed to skill, whereas losses are attributed to external factors and this pattern helps preserve overconfidence even when performance is weak (Barber & Odean, 2001).

Digital environments add another layer of complexity. At the same time, they create new possibilities for improving how people judge their own performance. Adaptive learning systems and decision support tools can now provide personalized feedback in real time. They track both performance and confidence across repeated decisions. This allows systems to adjust to each individual instead of offering the same feedback to everyone. Many of these tools are based on reinforcement learning principles and aim to improve calibration through repeated adjustments over time. Gradually, this process can help align confidence more closely with actual performance (Bansal et al., 2021).

However, technical sophistication alone is not enough. Trust plays a central role. If users perceive feedback as unclear, unfair, or overly harsh, they may disengage from the

system. In such cases, even accurate feedback can be ignored, which limits the system's ability to improve judgment.

Overall, feedback is a multidimensional mechanism for aligning confidence with performance. Its effectiveness depends on several factors, including immediacy, meaning how quickly feedback is given, clarity, meaning how easy it is to understand, and framing, meaning how the message is presented. Immediate and specific feedback strengthens the perceived connection between action and outcome. Delayed or ambiguous feedback weakens learning. Supportive delivery encourages reflection, whereas threatening feedback can reinforce bias. Importantly, feedback does not affect all forms of overconfidence in the same way. It is most effective in reducing overestimation, moderately effective for overplacement, and least effective for overprecision. This evidence suggests that feedback is a central but nuanced regulator of confidence. The following section examines these different forms of overconfidence in more detail and analyzes how they shape decision making outcomes (Moore & Healy, 2008).

2.4 Neuroscience of Confidence and Miscalibration

In addition to psychological and behavioral theories, recent research has examined confidence from a neuroscientific perspective. This line of work seeks to explain how confidence judgments are represented in the brain and updated through experience and feedback. By moving beyond observable behavior, neuroscience investigates how internal evaluation processes are implemented at the neural level. The available evidence supports the view that confidence is not merely a subjective feeling. It is a measurable cognitive process grounded in identifiable neural mechanisms (Fleming & Dolan, 2012).

Early studies in cognitive neuroscience demonstrated that confidence judgments rely on brain systems involved in decision making, error detection, and learning. When individuals make a decision, the brain generates not only a choice but also an internal estimate of how reliable that choice is. This estimate is closely linked to the strength and clarity of the underlying evidence. Stronger internal signals tend to produce higher confidence, whereas weaker or more ambiguous signals are associated with lower confidence.

One of the central regions involved in this process is the prefrontal cortex. Research using functional magnetic resonance imaging has shown that activity in this region correlates with subjective confidence, even when objective accuracy remains constant (Fleming et al., 2010). This pattern indicates that confidence reflects additional monitoring processes beyond basic task performance. In particular, the anterior prefrontal cortex appears to play a key role in metacognitive evaluation, defined as the ability to assess one's own cognitive performance.

Another critical region involved in confidence processing is the anterior cingulate cortex, which plays a central role in error detection and conflict monitoring. When outcomes fail to match expectations, activity in this region increases, indicating that behavioral adjustment may be required (Ridderinkhof et al., 2004). This response forms a key component of feedback processing. When an individual learns that a decision was incorrect, the brain generates a prediction error signal that supports the updating of future confidence judgments. If these signals are weak or absent, confidence levels tend to remain stable, even when they are inaccurate (Schultz, 2016).

Neuroscientific research also supports learning based models of confidence. Drawing on reinforcement learning theory, studies show that confidence updating follows principles similar to reward learning. Outcomes that exceed expectations tend to increase confidence, whereas outcomes that fall short should reduce it. In practice, however, this process is often

asymmetric and individuals typically adjust their confidence more strongly after positive outcomes than after negative ones. This imbalance contributes to persistent overconfidence (Lefebvre et al., 2017).

Importantly, neural evidence suggests that miscalibration can arise even when the underlying brain systems function normally. Overconfidence does not necessarily reflect a malfunction, but rather a systematic bias in how information is weighted and integrated over time. Feedback that is delayed, noisy, or emotionally threatening may produce weaker neural learning signals, thereby reducing its influence on adjustment. This helps explain why certain environments fail to correct miscalibration even when objective feedback is available.

Research also highlights substantial individual differences in these neural processes. Some individuals show stronger sensitivity to error signals, which is associated with better calibration over time. Others exhibit reduced neural responses to negative feedback. These individuals are therefore more susceptible to persistent overconfidence (Fleming & Lau, 2014). This pattern suggests that calibration depends not only on the availability of external feedback, but also on internal sensitivity to learning signals.

Overall, neuroscientific findings reinforce key insights from psychological and behavioral theories. Confidence reflects an ongoing interaction between evidence, feedback, and learning mechanisms in the brain. By identifying the neural systems involved, neuroscience provides an additional explanatory layer for understanding why feedback improves calibration in some contexts but not in others (Moore & Healy, 2008).

2.5 Practical Implications of Overconfidence and Feedback

The effects of overconfidence and feedback extend well beyond laboratory research. In everyday life and professional contexts, confidence judgments shape how people learn, make decisions, guide others, and assess risk. In fields such as healthcare, education, organizations, and financial markets, the ability to calibrate confidence accurately influences not only individual outcomes but also the performance and safety of larger systems. While overconfidence can motivate action and persistence, it can also lead to serious errors when certainty exceeds available evidence. Feedback plays a central role in linking perceived ability with actual performance, yet its effectiveness depends strongly on context, structure, and emotional response. Understanding these conditions is essential for translating theoretical findings into real world applications (Moore & Healy, 2008).

It is important to look at all of these separately. Healthcare, diagnostic overconfidence represents one of the most consequential forms of miscalibration. Physicians often rely on experience and pattern recognition, which supports rapid decision making but may also increase certainty beyond what the available evidence justifies. Evidence from Berner and Graber (2008) suggests that a substantial proportion of medical errors is associated with clinicians overestimating diagnostic accuracy. Feedback tools such as audits, simulation training, and peer review discussions can improve calibration by providing clear and specific information about diagnostic outcomes. These systems enable clinicians to compare their judgments with objective results and peer evaluations. However, medical environments are often shaped by hierarchy and intense time pressure, where admitting uncertainty may be perceived as weakness. Emotional framing therefore plays a central role in determining whether feedback leads to learning. When feedback is presented as developmental rather than punitive, clinicians are more likely to adjust their confidence. Institutions that encourage open

discussion of errors tend to develop stronger learning cultures and more accurate self assessment over time (Winne & Hadwin, 1998).

Education provides another domain in which confidence calibration is essential. Students must evaluate their own knowledge accurately in order to regulate learning effectively. Overconfident students may reduce study effort too early, whereas underconfident students may allocate effort inefficiently. Formative assessments, self testing, and reflective learning activities help students identify gaps between confidence and understanding. Research by Zimmerman and Schunk (2011) shows that continuous feedback improves both confidence accuracy and academic performance. Metacognitive training further encourages students to engage actively with feedback rather than avoid it. This approach strengthens calibration and resilience by framing mistakes as part of the learning process and many modern learning systems build on these principles by adjusting task difficulty and providing real time feedback. When feedback is perceived as supportive rather than judgmental, students engage more deeply and develop durable self evaluation skills.

Similar dynamics are visible within organizations. Overconfidence often appears in the form of overplacement, where individuals believe they perform better than their peers. This tendency is particularly relevant in leadership and performance evaluation. Leaders who consistently overrate their relative ability may underestimate risks, dismiss feedback, or rely excessively on their own judgment. At the same time, moderate confidence can inspire trust and motivation within teams and comparative feedback tools such as multisource evaluations and team based reviews provide leaders with broader perspectives on their behavior. Their effectiveness, however, depends largely on organizational culture. When feedback systems are perceived as punitive, individuals tend to reject negative input. When feedback is framed as developmental, they are more willing to recalibrate their self perception. Research shows that

organizations that promote psychological safety achieve more accurate calibration across hierarchical levels (Edmondson, 2019). This illustrates that feedback operates not only through information, but also through trust, relationships, and shared goals.

On the other side, in finance and entrepreneurship, overconfidence has mixed effects. Optimistic beliefs can support innovation, persistence, and risk taking. Many entrepreneurial successes depend on pursuing ideas that appear unlikely when judged by statistical expectations alone. At the same time, excessive confidence can distort risk evaluation and contribute to repeated failure. Barber and Odean (2001) showed that overconfident investors trade more frequently and earn lower returns. Similarly, Camerer and Lovallo (1999) found that entrepreneurs often overestimate their probability of success, following that confidence can therefore drive action, but without calibration it can also magnify error.

Feedback can help balance ambition and realism when it is used in the right way. It shows people where they stand. Feedback that includes information about how likely a prediction was helps decision makers see not only whether they were wrong, but also how far their estimate was from the actual outcome. In entrepreneurship, structured project reviews and post project evaluations encourage people to reflect on how they made decisions, not only on whether the result was good or bad. This change in focus supports learning over time. It improves judgment without lowering motivation. The goal is not to remove confidence completely, but to align optimism more closely with the available evidence.

As mentioned in previous chapters, digital technologies have further increased the relevance of feedback and confidence calibration, since data driven systems can now provide personalized feedback in healthcare, education, and workplace performance. These systems can collect confidence judgments and present calibration information that compares expected and actual outcomes. Early evidence suggests that such feedback improves judgment

accuracy and decision efficiency (Bansal et al., 2021). However, technical precision alone does not guarantee effectiveness. Users must trust the system and perceive feedback as supportive rather than evaluative. When feedback is experienced as critical, resistance may emerge, similar to patterns observed in interpersonal feedback and effective human technology interaction therefore depends on feedback loops that force human judgment rather than replace it.

Across different areas, some general principles explain when feedback improves calibration. Timing is important. Feedback should be given soon after a decision and it should be clear and specific. When people can easily connect their action to the outcome, learning is stronger (Stone, 1994).

Feedback should also fit the type of overconfidence involved. Overestimation improves most when people receive clear information about whether they were right or wrong, because this directly shows the gap between confidence and performance (Fischer & Budescu, 2005). Overplacement is different. Since it is based on comparison with others, it improves when people receive information about their relative standing, especially in a safe and non threatening setting (Pulford & Colman, 1997). Overprecision is often harder to change. Simple right or wrong feedback is usually not enough. People need clear information about uncertainty and probability to adjust their level of certainty (Soll & Klayman, 2004).

Emotional framing also matters, because supportive feedback encourages reflection and learning, meaning that feedback that feels critical or threatening can lead to defensiveness and reduce improvement. Since these elements work together and accurate information alone is not sufficient (Kluger & DeNisi, 1996).

Even though much research has been conducted, open questions remain. People differ in how they respond to feedback. Some adjust quickly, while others continue to hold inaccurate beliefs even after repeated correction. In addition, improvements in calibration may decrease once feedback is removed. This is sometimes called feedback decay and maintaining accurate calibration over time is therefore difficult (Koriat, 2012).

This leads to a broader question if perfect calibration is always the best choice. Moderate overconfidence can increase motivation and persistence. However, too much caution can limit action and decision making. The aim is not to remove confidence completely, but to keep it at a level that remains useful in uncertain situations.

2.6 Summary and Conceptual Framework

Across the literature reviewed, one clear conclusion appears. Confidence is not fixed, develops over time and is strongly influenced by feedback. From the early work of Meehl (1954) to more recent neuroscientific and computational models, confidence judgments are described as estimates that change through experience rather than stable personality traits. Feedback connects what people believe internally with what actually happens externally. Its impact depends on timing, clarity, context, and emotional framing. When these elements fit together, confidence can adjust and become more accurate. When they do not, bias often remains.

Overconfidence is not a single problem since it includes different tendencies that require different forms of correction. Outcome based feedback helps reduce overestimation because it shows the gap between expected and actual performance. Comparative feedback can reduce overplacement, especially when it is delivered in a psychologically safe setting.

Probabilistic feedback addresses overprecision by making uncertainty visible and encouraging more realistic judgments. Each type of feedback therefore works at a different level.

At the theoretical level, several models point in the same direction. Signal Detection Theory explains how evidence relates to certainty. Dual process models describe how intuitive confidence forms quickly and is later adjusted through more analytical thinking. Reinforcement learning models show how differences between expected and actual outcomes drive updating. Together, these perspectives describe confidence as a process that is shaped by feedback loops involving cognition, emotion, and environment. It is not static and it changes with experience.

Empirical research also shows that feedback does not affect everyone in the same way. Expertise, motivation, and personality influence how people respond. Experts may start with better calibration, but they can still become overly precise in familiar domains. Novices often begin with larger errors, but they improve quickly when feedback is clear and immediate. Traits such as being open increase receptiveness, while anxiety or ego protection may reduce it. Not to forget context of feedback matters as well because supportive and transparent environments promote learning, whereas hierarchical settings can limit it. Feedback is therefore both informational and social.

These insights form the basis of the conceptual framework guiding the present study. The framework assumes that performance feedback influences decision accuracy through its effect on confidence calibration. Each decision produces an initial confidence judgment, while feedback provides new information. Recalibration then adjusts future judgments and over time, repeated cycles determine whether confidence moves closer to objective accuracy or remains biased.

To summarize how different feedback types target specific forms of overconfidence, Table 1 combines evidence from Moore and Healy (2008) with studies by Stone (1994), Soll and Klayman (2004), and Fischer and Budescu (2005). The table shows how each feedback type operates and where improvement is expected.

Table 1 *Feedback Types and Their Primary Effects on Overconfidence Dimensions*

Feedback Type	Primary Target Bias	Mechanism of Action	Expected Calibration Effect
Outcome Feedback	Overestimation	Reveals discrepancy between perceived and actual performance	Strong reduction in overestimation
Comparative Feedback	Overplacement	Provides information about relative standing within a group	Moderate reduction in overplacement
Probabilistic / Uncertainty	Overprecision	Highlights mismatch between subjective certainty and statistical accuracy	Gradual reduction in excessive certainty

The synthesis provided in Table 1 suggests that feedback is most effective when it directly addresses the cognitive source of the bias. While outcome feedback provides a clear corrective signal for overestimation by revealing performance gaps, overprecision remains notably more resistant to change. This distinction is critical for the present research, as it highlights why simple performance results may not suffice to improve judgment in complex, uncertain environments. The framework also acknowledges interaction effects among overconfidence dimensions. Improvements in one domain do not automatically generalize to others and may generate secondary distortions if misinterpreted. Effective calibration therefore requires coordinated feedback strategies that address multiple mechanisms simultaneously.

Practically, this framework provides guidance for educational, organizational, and technological interventions. Educators can combine outcome feedback with reflective activities to improve learning accuracy. Organizations can implement comparative feedback systems that encourage reflection without threat. Designers of decision-support technologies can integrate uncertainty visualization tools to promote calibrated interpretation. Across contexts, feedback should be developmental rather than evaluative.

In summary, this framework describes confidence calibration as a continuous process in which people compare their beliefs with evidence and adjust their judgments based on their environment. Although bias is a normal part of human thinking, it can be reduced when feedback is timely, clear, and supportive. The framework brings together cognitive, emotional, and social factors to better explain how feedback influences decision-making.

Building on this synthesis, the present study empirically examines how distinct feedback types affect confidence calibration across overestimation and overplacement, while monitoring overprecision as a secondary outcome. By comparing outcome-based and comparative feedback within a single experimental design, the study aims to clarify which feedback mechanisms most effectively align confidence with accuracy, contributing both theoretical insight and practical guidance for improving judgment in feedback-rich environments.

3. Research Question and Hypotheses

3.1 Research Gap

Previous part showed that confidence is not a fixed personal attribute, but something that can change over time through experience and feedback. Many studies suggest that

feedback can improve calibration. However, the results are not always consistent and most existing research focuses either on outcome feedback, where participants learn whether their answer was correct or incorrect, or on comparative feedback, where individuals receive information about how they performed compared to others. Both types have been studied separately, but they are rarely directly compared within the same experimental setting. Due to this, it is still unclear whether outcome and comparative feedback influence confidence in the same way or whether they affect different forms of overconfidence. In addition, many studies treat calibration as one overall measure. They often do not clearly separate overestimation from overplacement.

If these dimensions are based on different psychological processes, it is reasonable to assume that they may also respond differently to different types of feedback. The present study addresses this gap by directly comparing outcome based and comparative feedback within one unified experimental design. While overprecision was discussed in the theoretical section, this study focuses specifically on overestimation and overplacement, as these are most relevant for the general knowledge task used in this experiment.

3.2 Research Question

Based on the literature review and the identified gap, the central research question of this study is:

Research Question:

RQ1: How do different feedback types influence confidence calibration across the dimensions of overestimation and overplacement?

3.3 Hypothesis Development

Based on the theoretical background and identified research gap, the following hypotheses are formulated.

H1: Outcome based feedback significantly reduces overestimation compared to comparative feedback and no feedback.

H2: Comparative feedback significantly reduces overplacement compared to outcome based feedback and no feedback.

H3: Participants who receive feedback show better overall confidence calibration than participants who receive no feedback.

These hypotheses reflect the assumption that different types of feedback influence different dimensions of overconfidence. Outcome based feedback is expected to improve individual accuracy judgments, while comparative feedback is expected to influence relative self evaluation. Overall, the presence of feedback is expected to improve alignment between confidence and actual performance.

3.4 Measurement of Variables

To evaluate calibration accuracy, objective performance scores will be compared with self-reported confidence ratings. Calibration will be operationalized as the absolute difference between mean confidence and mean accuracy. Smaller differences indicate better alignment

between subjective judgment and actual performance, while larger differences reflect stronger miscalibration.

In addition to overall calibration, two specific dimensions of overconfidence will be examined. Overestimation will be calculated as the difference between mean confidence and mean accuracy, where positive values indicate that participants overestimate their performance. Overplacement will be assessed by comparing participants' perceived percentile ranking with their actual percentile ranking within the sample, capturing inaccuracies in relative self-assessment.

By distinguishing between these components, the analysis allows for a more precise examination of how different feedback types influence specific dimensions of confidence calibration.

4. Experiment Design & Methodology

Given the scope of recruiting participants across different channels, data collection for this study was conducted through an online survey built in Google Forms. An online format made it possible to reach a broader and more varied sample, and it also simplified the administration of separate feedback conditions without participants becoming aware of the other versions. The study follows a between-subjects design. Each participant was randomly assigned to one of three groups and completed only one version of the survey. The core task was the same for everyone, but what happened after the quiz differed depending on the assigned condition.

4.1 Survey Structure

Each version of the survey followed the same four-stage structure. The content of Stage 3 varied by group, while everything else remained identical.

At the start of the survey, participants were given a brief description of the study. This was intentionally kept short so as not to prime them toward any particular response pattern. They were told the study looked at how people judge their own performance on knowledge tasks, and that all responses would remain anonymous. After reading this, participants filled in a short demographics section asking for their age group, gender, and highest level of education completed. These variables are useful for later analysis, particularly if patterns emerge across subgroups.

After completing the demographics section, participants moved to the main part of the survey. In the questionnaire itself, there were 10 trivia questions drawn from a range of topic areas including history, science, geography, and general culture. The questions were presented in a fixed order, the same across all three groups and after each question, participants indicated how confident they were in their answer using a table from 0 to 10 where 0 is lowest and 10 is highest. It was important to collect this rating immediately after each question rather than at the end, since confidence judgments made in the moment tend to be more reflective of actual perceived certainty than retrospective estimates made at the end..

Once the quiz was complete, participants moved to the feedback stage and what they saw here depended on which group they had been assigned to. This stage is described in detail in Section 4.2.

The final stage, in which each person was asked what percentage of other participants they thought they had outperformed, was the same for all participants. This question was

designed to capture overplacement without reference to the feedback they had just received. Participants in Groups A and B were also asked one additional item about how they experienced the feedback, rated on a scale from very unhelpful to very helpful. This was included partly as a manipulation check, but also because the literature suggests that how feedback is perceived emotionally can shape whether it actually influences subsequent confidence judgments.

4.2 Experimental Groups

Participants were assigned randomly to one of three groups by distributing separate survey links through different channels and important to mention is that they were not told that other versions of the survey existed.

Group A received outcome-based feedback. After finishing the quiz, each participant was shown their actual score, their estimated score, and the gap between the two. Point of this was to give them a direct and concrete signal of how far their self-assessment was from their actual performance. Research by Fischer and Budescu (2005) suggests this kind of direct accuracy feedback tends to reduce overestimation more than other feedback types, since it makes performance errors explicit rather than leaving them to be inferred.

Group B received comparative feedback. Participants in this group were shown their total score alongside an estimated rank relative to others who had taken the survey. Specifically, they were informed of the percentage of participants they scored higher than, giving them a direct signal of their relative standing. The purpose of this condition is to address overplacement specifically. Knowing whether one performed better or worse than the average is more relevant to reducing inflated relative self-assessments than knowing which questions

were answered incorrectly. Pulford and Colman (1997) note that comparative feedback can meaningfully shift overplacement, though the effect tends to be modest, particularly when participants feel the comparison is unfavorable.

Group C served as the control condition. These participants received no feedback at all after completing the quiz. They moved directly from the quiz to the post-quiz measures. Having a no-feedback group is important because it allows for a cleaner comparison because any difference in calibration between Groups A and B and this group can more plausibly be attributed to the feedback itself rather than simply to the process of reflecting on one's performance.

4.3 Variable Description

The key variables used in this study are summarized in Table 2 below. A brief explanation of each follows.

Table 2. Variable Overview

Variable type	Variable	Operationalization
Independent variable	Feedback type	Three conditions: outcome feedback, comparative feedback, no feedback (control)
Dependent variable	Overestimation	Mean confidence rating minus mean accuracy score (positive value=overconfident)
Dependent variable	Overplacement	Self-estimated percentile rank minus actual percentile rank within the sample
Controlled variable	Quiz content	Same 10 questions in the same order across all three groups
Controlled variable	Confidence scale	0 to 100 percent slider after each question, identical wording in all versions
Moderator variable	Prior knowledge	Self-reported familiarity with general knowledge topics, collected in demographics
Mediator variable	Feedback perception	Single item rating of how helpful the feedback felt, collected in Groups A and B only

The independent variable is feedback type, which takes one of three values depending on group assignment. The two main dependent variables are overestimation and overplacement. Overestimation is calculated by subtracting mean accuracy from mean confidence across the 10 questions. Positive result means the participant was, on average, more confident than their performance justified. Overplacement on other hand works differently because it compares the rank a participant believed they had achieved against their actual rank in the sample. In this case as well, a positive value reflects an inflated sense of relative performance.

Quiz content and the confidence scale are both treated as controlled variables. All participants saw the same 10 questions in the same order and used the same scale 0 to 10 to report their confidence. This was very important in order to make meaningful comparisons across groups.

Two further variables are included to account for factors that could complicate the picture. Prior knowledge of the quiz topics is treated as a moderator, because someone with stronger general knowledge might respond differently to outcome feedback than someone who struggled with the questions. It was assessed through a self-report item in the demographics section rather than through any objective measure, which is a limitation. Feedback perception in Groups A and B was noted as a potentially relevant factor, given that Kluger and DeNisi (1996) and Dunning (2011) both argue that feedback which feels critical or unfair may prompt defensiveness rather than recalibration. However, due to the scope of the study, it was not formally tested as a mediator and is acknowledged here as a limitation.

4.4 Participants and Recruitment

The study aimed to collect between 75 and 90 completed responses in total, with roughly 25 to 30 participants per group. This was considered a reasonable target given the timeline and recruitment channels available, and it should be sufficient to run between-group comparisons using one-way analysis of variance.

Recruitment was carried out through a combination of university mailing lists, social media, and personal contacts. Recruitment began with a single survey link shared broadly, and once enough responses had come in, the three separate group links were distributed individually by email to maintain separation between conditions. Each participant got their

results and feedback in addition and they were not told that multiple versions existed. No specific background or expertise was required to take part, and participation was voluntary.

To be included in the analysis, participants had to be at least 18 years old and complete all sections of the survey. Responses showing implausible patterns, for instance identical confidence ratings across all 10 questions, were flagged and excluded. Following the approach used by Moore and Healy (2008), any participant who did not provide a usable percentile estimate in the post-quiz section was excluded from the overplacement analysis, though their data was retained for the overestimation analysis where applicable.

5. Results

This chapter presents the findings from the two-round study. The structure follows the order of the hypotheses: descriptive statistics first, then assumption checks, then between-group comparisons, and finally within-group changes. All analyses are based on the 84 participants who completed both rounds 28 per group.

5.1 Descriptive Statistics

The final sample was perfectly balanced: 28 participants in each of the three groups. This equal distribution simplifies the interpretation of group comparisons and was not planned from the outset, but reflects the final state of the data after cleaning and matching.

At baseline, all three groups showed the same basic pattern. People thought they did better than they actually did. The mean estimated score was consistently higher than the actual score across all groups, and most participants placed themselves above the median in terms of percentile rank. This replicates the well-established overconfidence finding and provides a solid starting point for examining whether feedback changes anything.

The groups were not perfectly equivalent at baseline, which matters. Group A started with the largest overestimation gap ($M = 1.63$) and the highest overplacement gap ($M = 13.35\%$). Group C was the most modestly overconfident of the three (overestimation: $M = 0.89$; overplacement: $M = 2.80\%$). These differences were not caused by the intervention they existed before any feedback was given but they do mean Group A had the most room to improve. This should be kept in mind when reading the change scores. Table 3 shows the full Round 1 descriptive statistics.

Table 3. Descriptive Statistics Round 1

Variable	Group A (n=28)	Group B (n=28)	Group C (n=28)	Total (n=84)
Actual Score R1 — M (SD)	5.96 (1.60)	5.71 (1.21)	5.71 (1.70)	5.80 (1.50)
Estimated Score R1 — M (SD)	7.70 (1.66)	6.96 (1.81)	6.61 (1.64)	7.09 (1.74)
Overestimation Gap R1 — M (SD)	1.63 (2.39)	1.28 (2.07)	0.89 (2.22)	1.26 (2.23)
Est. Percentile % R1 — M (SD)	60.56 (26.69)	46.63 (24.94)	44.89 (21.02)	50.69 (25.04)
Actual Percentile % R1 — M (SD)	45.63 (27.93)	39.00 (27.03)	42.10 (30.26)	42.24 (28.23)
Overplacement Gap R1 — M (SD)	13.35 (42.14)	9.59 (38.70)	2.80 (32.41)	8.58 (37.75)

Note. M = Mean; SD = Standard Deviation. Overestimation Gap = Estimated Score minus Actual Score. Overplacement Gap = Estimated Percentile minus Actual Percentile. All values rounded to two decimal places.

Table 4 shows the Round 2 results and the change scores. The pattern is fairly clear. Group A moved the most on both dimensions. Their overestimation gap fell from 1.63 to 0.50, and their overplacement dropped by over 22 percentage points shifting from clear overplacement to slight underplacement on average. Group B showed smaller changes on both measures. Group C, with no feedback, barely moved on overestimation and actually became slightly more overplaced between rounds.

One additional finding is worth flagging before moving to the tests. Group A also improved their actual quiz score from $M = 5.96$ in Round 1 to $M = 6.96$ in Round 2. Neither Group B nor Group C showed a comparable improvement. Whether this reflects a learning effect, a slightly easier second quiz, or something else is discussed in Chapter 6.

Table 4. Descriptive Statistics Round 2 and Change Scores ($\Delta = R2 - R1$)

Variable	Group A (n=28)	Group B (n=28)	Group C (n=28)	Total (n=84)
Actual Score R2 — M (SD)	6.96 (1.20)	6.32 (1.68)	5.64 (1.99)	6.31 (1.72)
Estimated Score R2 — M (SD)	7.46 (1.62)	7.21 (1.57)	6.31 (1.81)	7.00 (1.72)
Overestimation Gap R2 — M (SD)	0.50 (1.60)	0.89 (1.64)	0.67 (2.60)	0.69 (1.98)
Est. Percentile % R2 — M (SD)	43.89 (26.24)	47.71 (26.16)	43.11 (22.41)	44.90 (24.78)
Actual Percentile % R2 — M (SD)	51.07 (23.61)	41.93 (27.24)	31.81 (29.26)	41.60 (27.64)
Overplacement Gap R2 — M (SD)	-7.18 (33.02)	5.78 (34.97)	11.30 (36.64)	3.30 (35.35)
Δ Overestimation — M (SD)	-1.15 (2.49)	-0.40 (2.20)	-0.22 (3.06)	-0.59 (2.62)
Δ Overplacement — M (SD)	-22.49 (57.30)	-4.28 (45.36)	+9.51 (44.84)	-5.75 (50.63)

Note. Δ Overestimation and Δ Overplacement represent the change from Round 1 to Round 2. Negative values indicate a reduction in overconfidence. All values rounded to two decimal places.

Figure 1: Mean Overestimation and Overplacement Gap — Round 1 vs Round 2

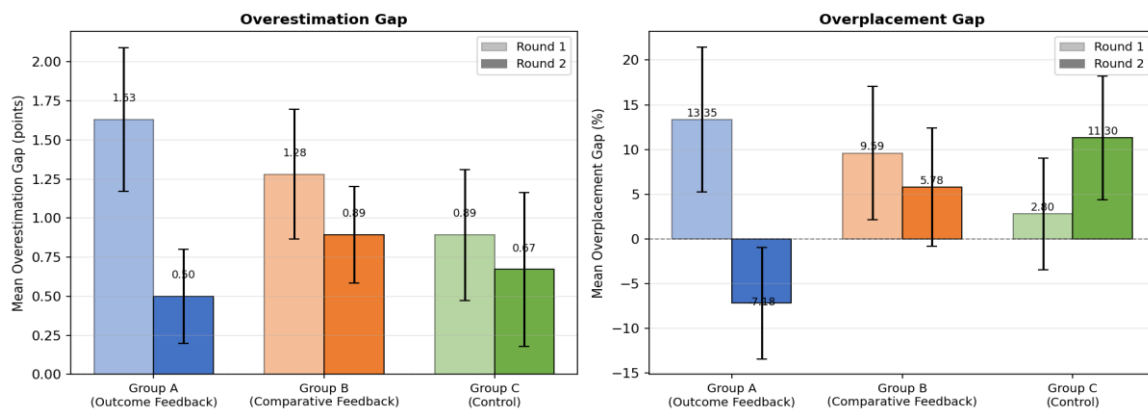


Figure 1. Mean overestimation gap (left) and overplacement gap (right) in Round 1 (lighter bars) and Round 2 (darker bars) by group. Error bars represent standard error of the mean.

5.2 Assumption Checks

5.2.1 Normality

Shapiro-Wilk tests were run on the change scores for each group. For the change in overestimation, Group A violated normality ($W = 0.870$, $p = .003$), while Groups B ($W = 0.924$, $p = .064$) and C ($W = 0.972$, $p = .643$) did not. This is not unusual, as Group A showed

considerable variation in change scores in both direction and because of this violation, both parametric and non-parametric tests are reported for overestimation to make sure the conclusions are robust.

For overplacement change, all three groups passed. Group A: $W = 0.963$, $p = .435$. Group B: $W = 0.981$, $p = .884$. Group C: $W = 0.959$, $p = .356$. No issues there.

5.2.2 Homogeneity of Variance

Levene's test showed that variance was comparable across groups for both variables. Overestimation change: $F(2, 81) = 1.76$, $p = .180$. Overplacement change: $F(2, 81) = 0.46$, $p = .636$. Both passed, so the standard ANOVA assumptions hold.

5.3 Between-Group Comparisons

To test whether the type of feedback made a difference, one-way ANOVAs were run on the change scores, with Kruskal-Wallis tests as a non-parametric backup. The goal was to see if the groups end up in different places after Round 2. Table 5 summarises the results.

Table 5. *Between-Group Inferential Tests on Change Scores*

Test	Variable	Group A	Group B	Group C	F / H / t	p-value	Effect Size
One-way ANOVA	Δ Overestimation	-1.15	-0.40	-0.22	$F=0.95$	$p=.390$	$\eta^2=.024$
One-way ANOVA	Δ Overplacement	-22.49	-4.28	+9.51	$F=2.84$	$p=.065$	$\eta^2=.068$
Kruskal-Wallis	Δ Overestimation	—	—	—	$H=1.25$	$p=.534$	—
Kruskal-Wallis	Δ Overplacement	—	—	—	$H=4.46$	$p=.108$	—
Post-hoc: A vs C	Δ Overplacement	—	—	—	$t=-2.29$	$p=.026$ ($p_{\text{bonf}}=.079$)	$d=-0.62$

Note. * $p < .05$. Effect sizes: η^2 for ANOVA; Cohen's d for post-hoc comparisons. Bonferroni correction applied to post-hoc tests (adjusted $\alpha = .017$).

5.3.1 Hypothesis 1 Overestimation

H1 predicted that Group A would show a greater reduction in overestimation than the other two groups. The ANOVA did not support this: $F(2, 77) = 0.95$, $p = .390$, $\eta^2 = .024$. The Kruskal-Wallis test agreed $H(2) = 1.25$, $p = .534$. No significant difference between groups.

The means do point in the right direction. Group A reduced overestimation the most ($M = -1.15$), then Group B ($M = -0.40$), then Group C ($M = -0.22$). But the differences are simply not large enough to reach significance given the sample size and within-group variability. H1 is not supported at the between-group level.

5.3.2 Hypothesis 2 Overplacement

H2 predicted that Group B would show the largest reduction in overplacement. This is where the results get interesting and a little unexpected. The ANOVA came close to significance: $F(2, 78) = 2.84$, $p = .065$, $\eta^2 = .068$. Not quite significant by the conventional threshold, but a medium effect size and a trend worth examining.

The pattern of means was not what H2 predicted. Group A dropped by 22.49 percentage points. Group B dropped only 4.28. And Group C went up by 9.51 points meaning participants with no feedback became more overplaced, not less. Post-hoc comparisons showed that the Group A vs. Group C difference was driving this pattern: $t = -2.29$, $p = .026$, $d = -0.62$. After Bonferroni correction the p-value rises to .079, just above the adjusted threshold, but the effect size is medium-to-large. No other pair differed significantly. H2 is not supported in its predicted direction, but the data do point to outcome-based feedback having a meaningful effect on overplacement just not comparative feedback.

5.3.3 Hypothesis 3 Overall Calibration

H3 predicted that receiving any feedback A or B would lead to better calibration than no feedback at all. The results offer partial support. Group A clearly moved toward better calibration on both dimensions. Group B did not its changes were small and non-significant across the board. Group C drifted slightly in the wrong direction. If H3 requires both feedback groups to outperform control, it does not hold. If the prediction is that at least one feedback condition makes a difference, the directional evidence supports it, even if the between-group tests do not all reach formal significance.

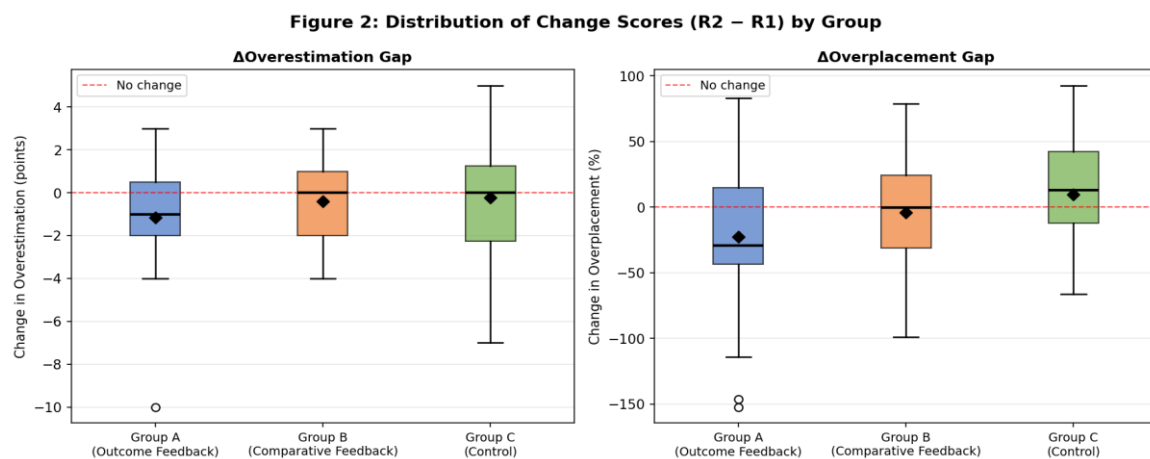


Figure 2. Distribution of change scores ($R2 - R1$) for overestimation (left) and overplacement (right) by group. Diamonds indicate group means. The dashed red line marks zero change. Negative values indicate reduced overconfidence.

5.4 Within-Group Paired Comparisons

Beyond the between-group picture, it is useful to ask whether each group changed significantly within itself. It is worth examining whether Group A actually improved from Round 1 to Round 2, and whether Group C stayed flat. Paired t-tests and Wilcoxon signed-rank tests were used to answer this and table 6 shows the full results.

Table 6. Within-Group Paired Comparisons (Round 1 vs Round 2)

Variable	Group	R1 Mean	R2 Mean	n	t / W (Wilcoxon)	p-value
Overestimation Gap	A	1.63	0.48	27	t=2.39 / W=69	p=.024* / p=.018*
Overestimation Gap	B	1.28	0.88	25	t=0.91 / W=57	p=.372 / p=.349
Overestimation Gap	C	0.89	0.67	28	t=0.38 / W=137	p=.705 / p=.698
Overplacement Gap	A	13.35	-9.14	27	t=2.04 / W=112	p=.052 / p=.064
Overplacement Gap	B	9.59	5.31	27	t=0.49 / W=167	p=.628 / p=.597
Overplacement Gap	C	2.80	12.31	27	t=-1.10 / W=145	p=.281 / p=.301
Actual Score	A	5.96	6.96	28	t=-2.53	p=.018*
Actual Score	B	5.71	6.32	28	t=-1.62	p=.118
Actual Score	C	5.71	5.64	28	t=0.16	p=.875

Note. * $p < .05$. Both parametric (t-test) and non-parametric (Wilcoxon W) statistics are reported. Cohen's d effect sizes for Group A: Overestimation $d = 0.55$; Overplacement $d = 0.58$; Actual Score $d = 0.67$.

5.4.1 Group A Outcome-Based Feedback

Group A is the only group that changed significantly across rounds. Three findings stand out.

First, overestimation dropped from $M = 1.63$ to $M = 0.48$. Both the t-test ($t(26) = 2.39$, $p = .024$) and the Wilcoxon test ($W = 69$, $p = .018$) confirmed this. Effect size: $d = 0.55$, moderate. Second, overplacement shifted from $M = 13.35\%$ to $M = -9.14\%$ a swing of over 22 percentage points. The t-test was borderline significant ($t(26) = 2.04$, $p = .052$) and the Wilcoxon test similarly close ($p = .064$). Neither crossed the $\alpha = .05$ threshold, but the direction is clear and the effect size is meaningful ($d = 0.58$). Third, actual quiz performance improved significantly, from $M = 5.96$ to $M = 6.96$: $t(27) = -2.53$, $p = .018$, $d = 0.67$. This was not expected and adds an interesting wrinkle to the interpretation.

5.4.2 Group B Comparative Feedback

Group B showed no significant change on any measure. Overestimation: 1.28 to 0.88 ($p = .372$). Overplacement: 9.59% to 5.31% ($p = .628$). Actual score: 5.71 to 6.32 ($p = .118$). The changes are small and inconsistent. Knowing where you ranked relative to others did not move the needle for this group.

5.4.3 Group C Control

The increase in overplacement from 2.80% to 12.31% did not reach significance ($p = .281$), but the size of the shift is worth noting. Without any corrective information, people became more confident about their relative standing in Round 2, not less.

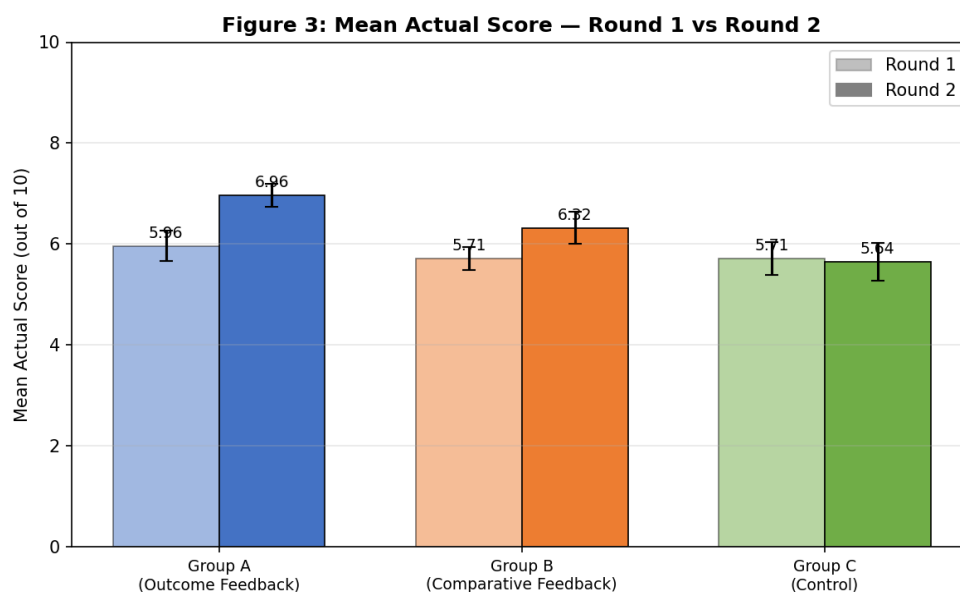


Figure 3. Mean actual quiz scores in Round 1 (lighter bars) and Round 2 (darker bars) by group. Error bars represent standard error of the mean.

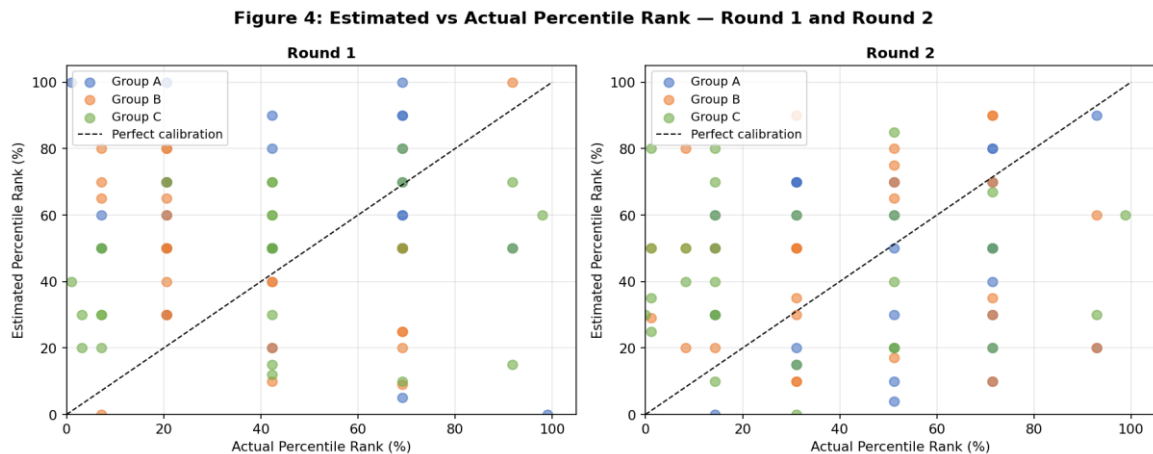


Figure 4. Estimated versus actual percentile rank for Round 1 (left) and Round 2 (right). The dashed diagonal represents perfect calibration. Points above the line indicate overplacement, points below indicate underplacement.

5.5 Summary of Findings

The clearest takeaway is that outcome-based feedback worked, while comparative feedback did not produce significant change in this study.

Group A reduced overestimation significantly, shifted overplacement by over 22 percentage points, and improved their actual quiz score. Group B showed no significant change on anything while group C, without any feedback, stayed flat on overestimation and drifted slightly toward greater overplacement.

H1 was not supported at the between-group level, though Group A did improve significantly within itself. H2 was not supported in its predicted direction. Group B did not outperform control, and the significant movement came from Group A instead. H3 received partial support since one feedback condition clearly outperformed no feedback, but the effect did not generalise to the other feedback condition. What this means theoretically and why comparative feedback may have fallen flat is addressed in Chapter 6.

6. Discussion

The aim of this study was to examine whether different types of feedback can reduce overconfidence, and whether outcome-based and comparative feedback affect different dimensions of it. The results show a fairly clear picture, even if not everything turned out as expected. This chapter discusses what the findings mean, why some hypotheses were supported and others were not, and what might explain the patterns that happened.

6.1 Outcome-Based Feedback and Overestimation

H1 predicted that Group A would reduce overestimation more than Groups B and C. At the between-group level this was not statistically confirmed, but the within-group results tell a different story. Group A was the only group to show a significant reduction in overestimation across rounds. The gap between what participants thought they scored and what they actually scored shrank meaningfully, and both the parametric and non-parametric tests agreed on this. With only 28 people per group the test simply did not have enough power to pick up the difference. Still the means went in the right direction, Group A improved most, Group B a little, Group C barely at all. The gap between them was too small to clear the significance bar.

The mechanism behind Group A's improvement seems straightforward. Participants were shown their actual score alongside what they had estimated. The gap was made explicit and concrete and it is hard to dismiss a number that is placed directly in front of you. This aligns with the argument made by Fischer and Budescu (2005), who found that direct accuracy feedback tends to reduce overestimation precisely because it removes ambiguity about how large the gap actually is. When people are left to draw their own conclusions, they often

underestimate how wrong they were. When the error is spelled out for them, they cannot ignore it.

Groups B and C showed much smaller changes. Group C barely moved at all, which is expected for a control group. Group B showed a slight reduction but nothing significant what suggests that knowing your rank among others does not by itself help people recalibrate how accurately they assessed their own performance. Relative standing and self-scoring accuracy are simply different things, and feedback targeting one does not necessarily transfer to the other.

6.2 Comparative Feedback and Overplacement

H2 was the most directly falsified hypothesis in the study. The prediction was that Group B, receiving information about their percentile rank, would reduce overplacement the most and that is not what happened. Group A showed the largest shift on overplacement, dropping by over 22 percentage points on average. Group B moved only 4 points while group C actually drifted upward by nearly 10 points. The one-way ANOVA for overplacement change was close to significant at $p = .065$, and the post-hoc comparison between Group A and Group C reached $p = .026$ with a medium to large effect size. The story emerging from the data is that outcome-based feedback affected overplacement too, even though it was not designed to do so specifically and comparative feedback did not.

Several explanations are worth considering to understand why Group B failed to reduce overplacement. The first is psychological, because people have a well-documented tendency to believe they are above average, a phenomenon sometimes called the better-than-average effect (Moore and Healy, 2008). When someone receives feedback telling them they ranked in

the 40th percentile, a common reaction is not recalibration but resistance. The information conflicts with a pre-existing self-image, and instead of updating that image, people find ways to question the feedback itself. Maybe the comparison group was unusually strong or maybe the questions were unfair or someone was using other sources to know the answers.

There is also a competitive dimension to this. Percentile rank is a social comparison because it tells you not just how well you did, but where you stand relative to others. For many people that framing triggers something closer to defensiveness than reflection. Knowing you are below the median does not feel like useful information but feels like losing. And when something feels like losing, the instinct is often to reject the outcome rather than learn from it. This may be especially true in a sample composed largely of personal contacts, as was the case in this study. Being told you underperformed compared to people you know personally adds a social sting that abstract comparative statistics in other studies do not carry.

A second explanation concerns the specificity of the feedback. Group A received very precise information. They saw their score, their estimate, and the exact gap between the two. Group B received a single number, a percentile rank, which is a more abstract and arguably more ambiguous piece of information because it does not tell you what you got wrong. It does not show you where your thinking broke down, but tells you where you ended up relative to others. For someone to use that information to actually recalibrate their confidence, they would need to make several inferential steps on their own. In practice, it seems most participants in Group B did not take those steps.

Pulford and Colman (1997) noted that comparative feedback can shift overplacement, but the effect tends to be modest and often depends on how the feedback is framed and received. If participants perceive the comparison as unfavorable, they are more likely to discount it and

this is consistent with what the data show here. Group B did not reject the feedback entirely, but they did not act on it either.

6.3 Group A and Overplacement

One of the more interesting findings in this study is that Group A showed the strongest movement on overplacement, even though the feedback they received was designed to address overestimation, not relative self-positioning. This was not predicted by H1 or H2. It deserves some explanation.

One possibility is that learning your exact score changes more than just your accuracy self-assessment. If you find out you scored 4 out of 10 when you thought you scored 7, you do not just update your estimate of your own score but you probably also update your sense of where you stand among others. The two are linked. A person who realises they performed much worse than expected is likely to revise their sense of relative standing downward as well, even without being given explicit rank information. Outcome feedback, in other words, may trigger a broader recalibration that goes into overplacement.

This interpretation is consistent with Moore and Healy's (2008) framework, which argues that overestimation and overplacement are related but distinct. When one is corrected through concrete performance information, the other can shift as a secondary consequence. Group A participants who saw that their self-estimates were far off may have concluded that they were not as knowledgeable as they had assumed, and that conclusion naturally extends to how they rank among others.

6.4 Group A and Quiz Performance

Group A also improved their actual quiz score significantly between rounds, from a mean of 5.96 to 6.96. Neither Group B nor Group C came close. This was not something the study set out to measure and should be interpreted carefully.

The most likely explanation is motivational because seeing that you scored lower than you expected creates a specific kind of pressure. That feeling of falling short of your own expectations probably made Group A engage more carefully in the second round, in a way that simply knowing your rank does not. Group B knew where they stood relative to others but not how far off their own self-assessment was, while Group C knew nothing at all. Without that concrete gap to react to, there was simply less to push against. It is also possible the second quiz was slightly easier and Group A happened to benefit more by chance, which cannot be ruled out without a counterbalanced design.

6.5 Comparative Feedback Performance

The failure of comparative feedback to produce significant change is worth examining more carefully, because it goes against a reasonable theoretical prediction. Social comparison theory, as developed by Festinger (1954) and extended in later work, suggests that people use information about others to calibrate their own self-assessments. If that is true, Group B should have updated their sense of relative standing when told where they ranked. They did not, at least not in any measurable way.

One factor that may have played a role is engagement with the feedback. The study did not include a direct measure of how carefully participants read and processed what they were shown. It is possible that some Group B participants glanced at their percentile rank without

really sitting with what it meant. Online surveys in general carry this risk because participants move through screens quickly, and a single number presented without context or explanation may not register as meaningful. Group A, by contrast, received a full breakdown of their answers, which is harder to skim and more likely to prompt genuine reflection.

Time is also a factor worth considering. Participants in this study completed both rounds within a limited timeframe, and there was no delay built into the design between receiving feedback and completing the second round. Research on feedback processing suggests that deeper recalibration often takes time. A percentile rank absorbed in thirty seconds during an online survey may simply not be enough exposure to shift a strongly held self-perception. Group A's feedback, being more detailed and more personally grounded, may have been easier to process quickly precisely because it was concrete rather than relational.

Finally, there is the question of emotional reception. Kluger and DeNisi (1996) argued that feedback which directs attention toward the self rather than toward the task tends to be less effective and can even trigger defensive processing. A percentile rank is almost entirely self-focused because it tells you nothing about the task itself but just tells you where you ended up in a social hierarchy. For participants who received an unflattering rank, this may have provoked a defensive response that actually reinforced their original self-assessment rather than correcting it. Outcome feedback, by keeping attention on the specific questions and answers, may have avoided this trap by staying task-focused.

6.6 Implications

The findings have a few practical implications worth mentioning, even if the study was not designed as an applied intervention. In educational and training contexts, where reducing

overconfidence is often a goal, the results suggest that showing people the exact gap between their performance and their self-assessment is more effective than simply telling them where they rank. Ranking information may feel threatening and therefore be discounted, while specific outcome information is harder to dismiss.

This has relevance for performance feedback systems in organisations as well. Many appraisal processes rely on comparative benchmarks, telling employees how they rate relative to peers or colleagues. The present findings suggest this may not be the most effective way to prompt recalibration. Feedback that directly links individual confidence or self-assessment to specific task outcomes is likely to be more useful, at least for improving accuracy of self-judgment.

It should be noted that these implications are tentative. The study used a general knowledge quiz with a student-adjacent sample, and the findings may not generalise straightforwardly to professional or high-stakes domains. Confidence in those contexts is shaped by years of experience and expertise in ways that a short online quiz cannot replicate. People do not stop resisting bad news about their ranking just because the context changes and they do not stop responding better to concrete personal feedback either.

6.7 Limitations

Several limitations of this study are worth acknowledging. The sample was recruited primarily through personal contacts and social networks. This introduces a degree of homogeneity that may limit generalisability. Participants were mostly from a similar age range and educational background, and being part of the researcher's personal network may have

introduced social desirability effects, particularly in Group B where comparative ranking could carry interpersonal meaning.

The study design did not include a time delay between feedback and the second round. This means it is not possible to assess whether any effects would persist over time or whether they reflect only immediate, surface-level recalibration. Confidence research often shows that short-term corrections fade without reinforcement (Lichtenstein and Fischhoff, 1977) and whether the changes observed in Group A would persist over days or weeks without further reinforcement cannot be determined from this design.

The quiz itself was the same in both rounds for all participants, which creates the possibility of practice effects contaminating the Round 2 scores. Group A's improvement in actual performance is the clearest example of this. Without using two different but equivalent quiz versions, there is no clean way to tell whether Group A improved because of the feedback or simply because they saw similar questions twice.

28 participants per group was enough to run the tests but not enough to detect smaller effects reliably. Several results pointed in the right direction but narrowly missed significance. A bigger sample would have given a clearer picture.

Finally, the study relied entirely on self-report for both confidence ratings and percentile estimates and no external validation of these measures was conducted. While this is standard practice in overconfidence research, it does mean the results are dependent on participants engaging honestly and thoughtfully with the survey.

7. Conclusion

This study compared three conditions to see whether the type of feedback people receive after a knowledge task changes how they assess themselves. Outcome-based feedback worked, comparative feedback did not while Group C stayed essentially flat. The hypotheses were partially supported at best, but the findings are clear enough to draw some meaningful conclusions and point toward what should come next.

7.1 Directions for Future Research

The most useful next step would be a replication with a larger and more diverse sample, a delayed follow-up measure, and counterbalanced quiz versions. That alone would resolve the three main methodological concerns at once and provide a more reliable test of whether the effects observed here hold beyond a single session.

It would also be worth testing a combined condition that delivers both outcome and comparative feedback together. Group A in this study received outcome information that incidentally shifted overplacement too. A condition that explicitly combines both might produce stronger effects on both dimensions and help clarify which mechanism is doing the work.

Why comparative feedback failed here is still an open question. The data point to several possible reasons, resistance, time pressure, and the abstract nature of a percentile rank, but cannot tell us which one actually drove it. A study that directly tests these mechanisms one by one would get us much closer to an answer.

Finally, the performance improvement in Group A is worth following up on. Outcome-based feedback might do more than just correct confidence. If it also helps people actually perform better on subsequent tasks, that matters well beyond the scope of this study and has real implications for how training is designed in education and professional settings.

7.2 Final Remarks

Overconfidence is not something wrong with person. It is simply how people tend to think about themselves when things are uncertain, and it shows up almost everywhere. The question was never really whether people are overconfident. The question is what, if anything, actually changes it and this thesis offers a small but concrete step toward that answer.

Outcome-based feedback, when delivered clearly and with specific numbers, can reduce overconfidence in a real and measurable way. Comparative feedback, at least in the way it was tested here, cannot. That difference matters for anyone who designs feedback systems, whether in schools, organisations, or healthcare. The details of how feedback is delivered are not a side issue. They are the whole point.

There is still much to understand. But the evidence from this study points in a consistent direction. Specificity works. Social comparison, at least on its own, does not.

References

- Adams, J. K., & Adams, P. A. (1961). Realism of confidence judgments. *Psychological Review*, 68(1), 33–45.
- Arkes, H. R., Dawes, R. M., & Christensen, C. (1986). Factors influencing the use of Bayesian strategies in prediction. *Organizational Behavior and Human Decision Processes*, 37(1), 93–120.
- Balla, J., & Heneghan, C. (2011). Overdiagnosis and overconfidence in medicine. *BMJ*, 342, d128.
- Bansal, G., Nushi, B., Kamar, E., Lasecki, W. S., Weld, D. S., & Horvitz, E. (2021). Improving human–AI collaboration by enhancing consistency in decision making. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1), 1–22.
- Barber, B. M., & Odean, T. (2001). Boys will be boys: Gender, overconfidence, and common stock investment. *The Quarterly Journal of Economics*, 116(1), 261–292.
- Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, 5(4), 323–370.
- Berner, E. S., & Graber, M. L. (2008). Overconfidence as a cause of diagnostic error in medicine. *The American Journal of Medicine*, 121(5 Suppl), S2–S23.
- Camerer, C., & Lovallo, D. (1999). Overconfidence and excess entry: An experimental approach. *American Economic Review*, 89(1), 306–318.
- Carminati, L. (2020). Behavioral economics: History, applications, and practical relevance. *Journal of Behavioral Economics for Policy*, 4(2), 5–15.
- Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science*, 243(4899), 1668–1674.
- Dunlosky, J., & Metcalfe, J. (2009). *Metacognition*. Sage.
- Dunning, D. (2011). The Dunning–Kruger effect: On being ignorant of one’s own ignorance. *Advances in Experimental Social Psychology*, 44, 247–296.
- Edmondson, A. C. (2019). *The fearless organization: Creating psychological safety in the workplace for learning, innovation, and growth*. Wiley.
- Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors*, 37(1), 32–64.
- Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, 7(2), 117–140.

Fischer, I., & Budescu, D. V. (2005). When do those who know more also know more about how much they know? *Organizational Behavior and Human Decision Processes*, 97(2), 135–155.

Flavell, J. H. (1979). Metacognition and cognitive monitoring. *American Psychologist*, 34(10), 906–911.

Fleming, S. M., & Dolan, R. J. (2012). The neural basis of metacognitive ability. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1594), 1338–1349.

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

Fleming, S. M., Weil, R. S., Nagy, Z., Dolan, R. J., & Rees, G. (2010). Relating introspective accuracy to individual differences in brain structure. *Science*, 329(5998), 1541–1543.

Glöckner, A., & Betsch, T. (2008). Multiple reason decision making based on automatic processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(5), 1055–1075.

Heine, S. J., Takata, T., & Lehman, D. R. (1999). Self-enhancement in Japan and the United States. *Journal of Personality and Social Psychology*, 72(6), 1244–1267.

Highhouse, S. (2008). Stubborn reliance on intuition and subjectivity in employee selection. *Industrial and Organizational Psychology*, 1(3), 333–342.

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.

Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise. *American Psychologist*, 64(6), 515–526.

Kepecs, A., & Mainen, Z. F. (2012). A computational framework for the study of confidence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1594), 1322–1337.

Klayman, J., Soll, J. B., González-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.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance. *Psychological Bulletin*, 119(2), 254–284.

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

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.

- Larrick, R. P. (2004). Debiasing. In D. J. Koehler & N. Harvey (Eds.), *Blackwell handbook of judgment and decision making* (pp. 316–338). Blackwell.
- Lefebvre, G., Lebreton, M., Meyniel, F., Bourgeois-Gironde, S., & Palminteri, S. (2017). Optimistic belief updating in the brain. *Nature Human Behaviour*, 1, 0067.
- 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.
- Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1982). Calibration of probabilities. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 306–334). Cambridge University Press.
- Lichtenstein, S., & Fischhoff, B. (1980). Training for calibration. *Organizational Behavior and Human Performance*, 26(2), 149–171.
- Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Lawrence Erlbaum.
- Masic, I. (2022). Decision making in healthcare. *Acta Informatica Medica*, 30(1), 65–72.
- Meehl, P. E. (1954). *Clinical versus statistical prediction*. University of Minnesota Press.
- Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review*, 115(2), 502–517.
- Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 26, pp. 125–173). Academic Press.
- Odean, T. (1998). Volume, volatility, price, and profit when all traders are above average. *Journal of Finance*, 53(6), 1887–1934.
- Pulford, B. D., & Colman, A. M. (1997). Overconfidence: Feedback and item difficulty effects. *Personality and Individual Differences*, 23(1), 125–133.
- Ridderinkhof, K. R., Ullsperger, M., Crone, E. A., & Nieuwenhuis, S. (2004). The role of the medial frontal cortex in cognitive control. *Science*, 306(5695), 443–447.
- Schultz, W. (2016). Dopamine reward prediction error coding. *Dialogues in Clinical Neuroscience*, 18(1), 23–32.
- Simon, H. A. (1957). *Models of man: Social and rational*. Wiley.
- Soll, J. B., & Klayman, J. (2004). Overconfidence in interval estimates. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(2), 299–314.
- Stewart, T. R. (2001). Improving reliability of judgmental forecasts. In J. S. Armstrong (Ed.), *Principles of forecasting* (pp. 81–106). Kluwer Academic.

Stone, E. R. (1994). Overconfidence in judgment. *Organizational Behavior and Human Decision Processes*, 58(2), 237–260.

Tenney, E. R., Logg, J. M., & Moore, D. A. (2008). The role of intuition in confidence judgments. *Journal of Experimental Psychology: General*, 137(3), 547–565.

Thompson, V. A., Prowse Turner, J. A., & Pennycook, G. (2011). Intuition, reason, and metacognition. *Cognitive Psychology*, 63(3), 107–140.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124–1131.

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277–304). Erlbaum.

Zimmerman, B. J., & Schunk, D. H. (2011). *Handbook of self-regulation of learning and performance*. Routledge.