The effects of mindfulness on the halo bias in performance reviews

Daniel Alács

Dissertation written under the supervision of Cristina Mendonça

Dissertation submitted in partial fulfilment of requirements for the MSc in International Management, at Universidade Católica Portuguesa and for the MSc in Strategy, Innovation and Management Control at WU University of Economics and Business, 05.09.2025.

Abstract

Title: The effects of mindfulness on the halo bias in performance reviews

Author: Daniel Alács

Assessments and appraisals serve as the basis for promotions, salary increases, training and developmental plans, as well as decisions about whose contract to terminate if need be. However, appraisals are prone to being influenced by biases, most importantly by the halo bias, potentially leading to suboptimal strategic choices. At the same time, loving-kindness meditation and mindfulness have been shown to remedy many biases. To understand how mindfulness induced through loving-kindness meditation can impact halo bias, an experimental study was conducted in the scope of this thesis.

The study examined the effect of a brief induction on mindfulness related measures and affect related measures. Furthermore, it aimed to assess whether people who have undergone a mindfulness induction through loving-kindness meditation exhibited lower levels of halo bias. The study revealed that participants exhibited higher levels of positive affect and significantly higher levels of state mindfulness. Most importantly, it showed that in the condition where halo bias is more likely to occur (moral description), decentering counteracted halo bias and significantly reduced it.

The study's results have implications for organisations, as short form meditation leads to significant increases in mood and mindfulness, which in turn might lead to higher engagement and satisfaction. Furthermore, it shows the potential of loving-kindness meditation as a tool to reduce halo bias.

Keywords: Loving-Kindness, Meditation, Mindfulness, Performance Appraisal, Performance Assessment, Performance Management

Summário

Título: Os efeitos da atenção plena no viés de halo nas avaliações de desempenho

Autor: Daniel Alács

As avaliações e apreciações servem de base para promoções, aumentos salariais, bem como para decisões sobre quais contratos rescindir, se necessário. No entanto, as apreciações são suscetíveis de serem influenciadas por vieses, principalmente pelo viés de halo, o que pode levar a escolhas estratégicas subótimas. Ao mesmo tempo, a meditação da bondade amorosa e a atenção plena têm demonstrado remediar muitos vieses. Para compreender como a atenção plena induzida pela meditação da bondade amorosa pode afetar o viés de halo, foi realizado um estudo experimental.

O estudo examinou o efeito de uma breve indução em medidas relacionadas com a atenção plena e medidas relacionadas com o afeto. Além disso, teve como objetivo avaliar se as pessoas que passaram por uma indução de atenção plena por meio da meditação da bondade amorosa apresentavam níveis mais baixos de viés de halo. O estudo revelou que os participantes apresentaram níveis mais altos de afeto positivo e níveis significativamente mais altos de atenção plena. Mais importante ainda, mostrou que na condição em que o viés de halo é mais provável de ocorrer (descrição moral), o descentramento reduziu significativamente o viés de halo.

Os resultados do estudo têm implicações para as organizações, uma vez que meditações curtas levam a aumentos significativos no humor e na atenção plena, o que, por sua vez, pode levar a um maior envolvimento e satisfação. Além disso, mostra o potencial da bondade amorosa para reduzir o viés de halo.

Palavras-chave: Bondade Amorosa, Meditação, Atenção Plena, Avaliação de Desempenho, Gestão de Desempenho

Glossary

α	Cronbach's index of reliability
SE	Standard Error
SD	Standard Deviation
M	Mean
MSA	Measure of sampling adequacy
Coeff	Coefficient
p	p-value
CI	Confidence Interval
df	Degrees of freedom
KMO	Kaiser-Meyer-Olkin
TMS	Toronto Mindfulness Scale
PANAS-SF	Positive Affect Negative Affect Scale – Short Form

Table of Contents

List of Figures

List of Tables

1 Introduction

> "*When you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind: it may be the beginning of knowledge, but you have scarcely, in your thoughts, advanced to the stage of science, whatever the matter may be.*" - Lord Kelvin (as cited in Houser, 2022)

In the opening quote, Lord Kelvin describes that if one cannot measure something, there is no way to improve it or have sufficient knowledge of what one is addressing or trying to address. This represents one of the many reasons why performance reviews, measurements, feedback sessions, and other tools meant to track one's activities are ubiquitous. Although researchers have been discussing the sensibility of performance reviews for a long time (DeNisi & Murphy, 2017), they are an indispensable part of work-life in many firms, and it is improbable that they will disappear anytime soon (Adler et al., 2016). Even the firms that may abandon performance reviews do not abandon the general notion of rating people, still measuring their work output and performance on several dimensions (Buckingham & Goodall, 2015). The reasons for questioning the usefulness of performance reviews include variations in ratings (Thorndike, 1920; Newcomb, 1929), the existence of biases (Tversky & Kahneman, 1974), the reviews' misalignment with organisational goals (Cleveland et al., 1989), and many other reasons. There have been several attempts to improve the efficacy of performance reviews (e.g.: behaviourally anchored scales, rater training), albeit with varying degrees of success (DeNisi & Murphy, 2017).

Ultimately, it is essential to acknowledge that the general goal of performance appraisals is to enhance the performance of individuals within the organisation and to thereby increase the effectiveness and success of the organisation (Randall & Sim, 2013). While performance reviews have been questioned regarding their efficacy and sensibility, another tool that could potentially counteract the drawbacks of reviews has gained significant popularity. This tool is mindfulness. Mindfulness works by fostering a more focused mind and maintaining a present focus (Singh et al., 2003).

As one of the significant drawbacks of performance reviews is the effect of biases on ratings, such as halo bias, recency effects, and spill-over effects (Hoyt, 2000), and bias reduction being one of the

attributes of meditation that has found support in the scientific literature (Maymin & Langer, 2021), it appears to be a logical step to try to combine these two tools to create a better and more accurate way of rating people's performance. This thesis proposes that mindfulness can serve as a tool to decrease biases when making judgments about another person. As such, this thesis aims to research the impact of mindfulness induced through meditation and its effects on performance reviews and biases. This master's thesis aims to contribute to the research on mindfulness and performance reviews. Combining these two research areas, this thesis explores the possibility of reducing biases in performance ratings through mindfulness, thereby creating a way to mitigate the inherent flaws of performance reviews. It sets out to find answers to the following research questions: RQ1: How can meditation influence reviews and feedback to facilitate employee growth? RQ2: How can loving-kindness meditation improve and make performance appraisals more accurate?

Although meditation in the western world has found its applications initially in psychotherapy (Kabat-Zinn et al., 1992), it quickly moved beyond this clinical setting. It has increased in popularity over the past few decades, as evidenced by its rising score on Google Trends (Google, n.d.). This trend is also reflected in the success of Headspace, an application that helps users sleep better and live a more stress-free life through meditation and mindfulness (Headspace, n.d.). Headspace boasts over ten million downloads in the Google Play Store alone (Headspace Inc., 2025). The popularity of mindfulness is no surprise as it can enhance mood (Hafenbrack et al., 2020), lead to better communication (Burgoon et al., 2002), as well as improved health (Kabat-Zinn, 1990). It appears to be a comprehensive solution to the various issues and challenges that organisations and their members encounter daily. Mindfulness and meditation, derived initially from the Buddhist religion, are often practised in a secularised fashion nowadays (Kucinskas, 2014). Due to the world becoming increasingly fast-paced and stressful, it seems natural that people have been increasingly interested in topics such as mindfulness and meditation that can reduce stress (Ospina, 2008).

Numerous entrepreneurs and public figures, including Steve Jobs (Gelles, 2015) and Bill Gates (Gates, 2018), have strongly advocated meditation and mindfulness. Several companies have started offering their employees mindfulness training and mindfulness retreats and creating working environments that allow people to take uninterrupted (meditation) breaks (Gelles, 2015). Prominently, Google offers several courses, such as *Search Inside Yourself*, programs like gPause, and retreat areas within their offices for their employees (Parcerisa, 2019). Google reasons that these programs help their employees relieve stress, communicate more effectively, and create an improved working environment (Parcerisa, 2019).

Problems that may arise through the practice of meditation are scarcely documented. Even though there have been cases of adverse effects of meditation (e.g. anxiety, depression, cognitive anomalies) (Farias, 2020), it may also be the case that meditation as a tool is still underutilised. The positive effects of mindfulness and meditation combined with the dissatisfaction of people with their work (Azeem et al., 2020) highlight the relevance of this topic. People at work feel stressed, overwhelmed, and anxious (Statista, n.d.) and in general often form judgments in a biased way (Tversky & Kahneman, 1974). They often experience the same biases repeatedly (Pohl & Hell, 1996; Raoelison & De Neys, 2019), often overinvest resources, fail to cut their losses, and do not always consider the most plausible and useful pieces of information (Gino, 2008; Odean, 1998; Ross & Staw, 1993).

This thesis is structured as follows. Chapter 1 represents this section and acts as an introduction to the topic. In chapter 2 the existing literature is reviewed focusing on performance management, performance appraisals as well as mindfulness and forms of meditation. Chapter 3 highlights the demographics of the participants, recruiting as well as the methodology of the experiment. Chapter 4 presents the results of the analysis whereas chapter 5 discusses the results, implications and limitations of the study. Lastly, chapter 6 contains a conclusion, which highlights the most important findings.

2 Literature Review

In the following sections the literature on performance management, appraisal and their shortcomings are reviewed. The literature on meditation, loving-kindness meditation and mindfulness is also examined in subsequent sections.

2.1 Performance Management

Performance management can be viewed as an interplay of a range of actions taken to enhance the performance of individuals, groups, and ultimately to increase organisational effectiveness (DeNisi, 2000). However, by including other definitions of this topic, the scope and understanding of performance management are expanded. Armstrong (2006) emphasises that organisational enhancement is achieved through managing performance within agreed frameworks, setting goals, creating and establishing shared standards and requirements for qualifications in order to make organisational success more probable in the short and long run. Whereas Lockett (1992), for example, emphasises fostering commitment towards the organisation, Walters (1995) expresses the supporting role of the organisation. Even though the definitions of performance management do not agree on all aspects, they do, however, agree in one aspect: performance management is focused on increasing organisational success through effective management of the business and its people. Therefore, performance management is a way for a firm to ensure the continuous improvement of its members by aligning the objectives of individuals, groups, and the firm as a whole. Furthermore, alignment between the firm's core values and its actions is of utmost importance (Zhao et al., 2016). By creating a shared vision, which all actors can support and align themselves with, organisational performance can be managed more effectively (Fletcher, 1993).

According to Armstrong (2006), performance management requires five main elements: agreement, measurement, feedback, positive reinforcement, and dialogue. These elements are essential to recognise possible areas for enhancement or growth in individuals and set goals that align with organisational needs (Armstrong, 2006). Additionally, these goals and growth areas are further explored, documented, and measured to track and compare performance, ensuring growth (Boswell & Boudreau, 2002; Randall & Sim, 2013). Lastly, constant feedback and performance-based reward systems that act as reinforcement ensure a periodic dialogue between parties (Cappelli & Conyon, 2017). This way, the behaviours, skills, and actions of people can be appropriately managed to fit organisational expectations.

As is apparent, performance management relies heavily on interpersonal relationships. Therefore, the notion of fair treatment is necessary to sustain dialogue and culture, which is supported by Winstanley and Stuart-Smith's (1996) ethical principles for performance management. These are centred around reciprocal respect for individuals, coupled with transparent and fair procedures and decision-making processes. These principles also emphasise the need and want of employees for guidance rather than control (Egan, 1995), which is why organisational support is crucial for development.

Furthermore, it is necessary to mention that these elements/processes are iterative and must be regarded as ongoing. Performance management requires managers, employees and other actors to assume responsibility for continuous progress and enhancement of skills, competences and processes (Armstrong & Baron, 2005). Additionally, performance management ideally also distinguishes between actions and results of those actions (Brumbach, 1988) to allow for a more nuanced and holistic picture of an individual's performance. As the relationship between a firm and an employee is not only a work contract, but also a psychological contract, distinctions between actions and results, fairness, trust and transparency carry great importance. A psychological contract in this sense refers to a belief system that is shared between employee and firm and entails expectations directed towards employees from the firm and vice versa (Guest et al., 1996). A positively perceived psychological contract can greatly enhance employees' alignment and identification with the company, and ultimately their satisfaction at work (Guest et al., 1996).

The performance management process follows four steps: planning, acting, monitoring, and reviewing (Deming, 1986). In this case, planning is a form of agreement that is ideally reached jointly between the employee and supervisor. Acting refers to the creation of a development plan and discussing specific actions that need to be taken to enhance performance. Monitoring refers to measurement and is not only a key step but also one of the most crucial aspects of performance management, allowing performance to be tracked to make comparisons and capture individual and organisational-level progress. Finally, the review process refers to reviews of performance, which have been touched upon in the five main elements of performance management as proposed by Armstrong (2006). More precisely, reviews are encompassed by ongoing feedback and dialogue. The four-step model is widely used and taught and is also known as the PDCA (Plan, Do, Check, Act) or Deming Cycle (Deming, 2018). The reason for its widespread use, mainly in quality management, but also in other fields, is that it has shown to be an efficient and simple tool to decrease preventable and frequently recurring issues (Realyvásquez-Vargas et al., 2018; Patel & Deshpande, 2017).

2.2 Performance Appraisal

Performance appraisals are a necessary part of managing performance, offering a systematic way to evaluate work-related behaviour, provide feedback, and review performance over a set period (Randall & Sim, 2013). A performance appraisal typically consists of a manager formally assessing an employee's performance in a top-down manner, usually on an annual basis (Armstrong & Baron, 2005). Interest in performance appraisals can be traced back to ancient China and has been part of human culture all throughout, manifesting in merit-based ratings (Murphy & Cleveland, 1995), military rankings (Cattell, 1906) or variations of scales as a non-replaceable part of research (Weiss, 1933). Over the years, several tools have been developed to streamline and improve rating efficiency. These are usually derivations of scales, but also include quota systems, forced rankings, distributions, oral and written assessments (Armstrong, 2006). Performance appraisals are an integral part of operations in companies. A survey by the Aberdeen Group (2010) reported that an estimated 91% of employers worldwide used appraisals. Although there is a substantial body of literature on scales, biases, and other aspects of performance management, there is less literature on appraisals from a management standpoint (Prendergast, 1999). This might also be because jobs and appraisals are highly complex, and performance itself is not easy to define, let alone measure (Lazear, 2000). Appraisals may be beneficial in the absence of metrics that are easy to measure (Baker et al., 1994). This is also supported by the notion that jobs involving more complex tasks typically more frequently include performance appraisals than jobs that involve simpler tasks (Brown & Heywood, 2005). The complexity of the job and the presence of the appraisals are also likely to be linked to merit-based pay (Brown & Heywood, 2005). Even in jobs where some aspects of an employee's performance are easier to measure, like in trading firms (portfolio performance), subjective measures play a significant role (Fenton-O'Creevy et al., 2005), which further underlines the importance of appraisals in complex job settings. However, it is also worth noting that performance appraisal as a whole has evolved into something broader than a topic solely about psychometric and evaluative issues (Fletcher, 2001; Levy & Williams, 2004). Performance appraisal encompasses not only evaluative components but also plays a key role in developmental areas, for example, identifying individual strengths and weaknesses and creating plans based on the identified areas for training and appropriate goal setting (Boswell & Boudreau, 2002).

Performance, as mentioned before, is complex to define. It can roughly be categorised into two categories, namely, task and contextual performance (Motowidlo & Scotter, 1994). Task performance encompasses metrics tied to the quality of the performed task e.g. structuredness of code. Contextual

performance encompasses performance beyond distinct tasks and includes a person's ability to create an atmosphere where effective and efficient work can be accomplished. The relationship between the rater and the ratee, specifically whether they are at the same hierarchical level, also influences the emphasis placed on one of the two performance types. Conway (1999) found that people on the same level pay more attention to interpersonal behaviours while superiors pay more attention to task-related performance. Contextual performance also relates to the notion of organisational spontaneity (George & Brief, 1992), which describes the key role of spreading good mood and spontaneous actions that foster goodwill for organisational success. Supporting others and exhibiting this form of spontaneity can be a sign of emotional intelligence (George & Brief, 1992). Furthermore, emotional intelligence is associated with self-awareness (Bratton et al., 2011), where people exhibiting this trait are often well-performing, positive members of a working environment (Sivanathan & Cynthia, 2002; Young & Dulewicz, 2007).

Additionally, depending on whether one is a well-performing or weak-performing individual, different organisational environments are more or less appealing. Recognition of individual contributions (Menefee & Murphy, 2004), bonus schemes that favour individual performance and opportunities for fast growth (Trank et al., 2002) are more desirable for high-performing individuals. These aspects, especially payment and its growth, have been tied to turnover rates. If the growth is low, high performers are more likely to leave the organisation (Allen & Griffeth, 2001), while low performers exhibit a lower turnover rate when salary does not greatly differ based on performance (Harrison et al., 1996). This, among many other aspects, showcases how important good appraisal and performance management systems can be for the overall effectiveness of an organisation.

The complexity of performance appraisals becomes further apparent when looking at the factors that influence them. As defined by Landy & Farr (1980), appraisals are influenced by different spheres, the sphere of the rater and ratee, the tool which is used to rate (which scale or other instrument), the reason for the rating within the organisational setting, and lastly the process itself, encompassing strategies of raters, ratees and bureaucratic constraints.

2.2.1 Rater Biases

Although there are numerous ways to measure and rate people, performance appraisals seem to have a myriad of issues. Even though performance management seems like an indispensable part of leading a team, managing an organisation, as well as staying competitive, people have been voicing their displeasure with appraisals and performance management for nearly as long as they exist (Kneeland, 1929; Pulakos & O'Leary, 2011). Thorndike (1920), with his famous 1920s paper, called attention to a constant error in psychological ratings. This not only sparked subsequent research (DeNisi & Murphy, 2017) but also brought attention to one of the most fundamental issues with ratings, namely rater biases. The situation concerning rater biases, which are at the core of performance ratings, have not changed over the last 100-plus years. Even before Thorndike, researchers like Edgeworth (1890) pointed out that receiving a good rating or passing an exam successfully not only depends on the examinee or ratee, but also considerably on the examiner/rater. This variability in scores and ratings is due to the fact that people are prone to biases, they have different levels of competence, frames of reference, and are influenced in their ratings by an interplay of things that can range from mood to automatic thought processes in the subconscious (Lumley & McNamara, 1995). To understand how significant this can be, the level of strictness of a judge can influence the variability of scores to the same degree as the ability of examinees (Cason & Cason, 1984). In general, these errors include restriction of range errors, which encompass leniency/severity (see Kneeland, 1929), central tendency errors, and halo errors (Dunnette & Borman, 1979).

The halo error, as Thorndike found, describes a statistically higher intercorrelation of items than expected for independent items like, for example, intelligence and physical qualities (Thorndike, 1920). This error, though identified by Wells (1907), was officially recognised in the research community and labelled as the halo error only after Thorndike's article. Later, Newcomb (1929; 1931) further supported the notion of this error through observations made from ratings of camp counsellors. Newcomb (1929) describes this error as inevitable and states that the main reason for this error lies in human logic of categorisation. For example, behaviours that can be logically linked under a single trait allow raters to make derivations that stem from their own logic rather than the actual behaviour of people. Thus, ratings are based on overall impressions, cognition and thought processes and are not as nuanced as intended based on the designs of tools, which include distinct items. Though the halo error seems ubiquitous, it seems to occur in differing degrees. On the one hand, it has been shown that halo effects are more prevalent when a superior is rating someone as opposed to people rating themselves (Heneman, 1974; Holzbach, 1978; Klimoski & London, 1974; Lawler, 1967). On the

other hand, ratings by peers are as biased as ratings from superiors (Lawler, 1967). Schneier (1977) also found that, depending on a person's ability to articulate and discriminate between differing dimensions that are indicative of ratee behaviour, halo error differs. The interplay of articulation and discrimination is referred to as cognitive complexity (Bieri, 1966). People with higher cognitive complexity are less prone to halo errors, whereas people with lower cognitive complexity are more prone to halo errors (Schneier, 1977). This suggests that the use of language and how one forms trains of thought and categorises attributes could potentially be used to decrease halo errors.

Leniency and severity errors describe the inclination of raters to either give higher or lower ratings on average (Holzbach, 1978). Leniency bias might also be a form of avoidance of discomfort on the raters' side and a form of motivated action, which depends on their personality and the context of the appraisal (Spence & Keeping, 2011). This means that raters will steer clear of ratings that do not align with goals and that cause undesired results (Spence & Keeping, 2011). Usually, these errors are identifiable when different rater groups have different average tendencies regarding ratings of the same ratee group (Sharon & Bartlett, 1969). This is also noticeable in data from firms, where systematic differences in evaluations are discernible depending on supervisors, with some giving higher ratings on average (Hall & Madigan, 2000). Though it must be pointed out that leniency errors, as opposed to halo errors, are more observable in self-ratings than ratings by supervisors or people on the same hierarchical level (Klimoski & London, 1974; Thornton, 1968). Leniency also goes hand in hand with a type of inflation of ratings, which can be due to differences in the interpretation of a scale (Hoyt, 2000). For example, for one supervisor, a score of four out of seven may be considered average. For another, it might be bad, though, numerically speaking, it is an above-average value. More concerning might be the overall perception of average scores, as the words satisfactory or meeting standards are perceived as highly negative by ratees (Perry & Porter, 1981). These differences in perception and understanding can lead to distortions in the mean (Hoyt, 2000).

One additional factor which significantly influences rater biases is the relationship between rater and ratee. Given that the rater/superior has taken a liking to the ratee, rater biases during appraisals are exhibited to a greater degree and superiors are also less likely to punish performance that falls below expectations (Lefkowitz, 2000). This positive effect may be due to similarities between rater and ratee (Wayne & Liden, 1995). There is also support suggesting that employees may foster liking through attempts to influence others' perceptions of them (Wayne & Ferris, 1990) and other tactics, which can be categorised along two dimensions: assertive-defensive and tactical-strategic (short-term vs. long-

term; Tadeschi & Melburg, 1984). Behaviours can include apologies, disclaimers, self-promotion, or attempts to enhance one's reputation (Ferris et al., 1994).

Other errors, biases and heuristics that may further lead to distorted and inaccurate ratings are the availability heuristic, anchoring bias, and overconfidence bias. The availability heuristic describes a mental shortcut, first explored by Tversky & Kahneman (1973), by which a person categorises or assesses the likelihood of an event based on its saliency or availability in their mind. This could lead to a person's assessment being overly influenced by past behaviour or even expectations towards the person. Coupled with overconfidence, which describes a form of mental bias which leads to a systematic overestimation of skills (Harvey, 1997; Lichtenstein et al., 1977; Wright, 1982) (e.g.: being an accurate judge of character) this may further lead to faulty judgment by disregarding outside opinion. This can be due to assuming one's own opinions and thoughts as superior to those of others, which can lead to the mentioned discounting of outside information (Moore & Healy, 2008).

Lastly, anchoring biases can lead to inaccurate assessments of people based on initial perception. The bias itself refers to the phenomenon where initially received/perceived information serves as an anchor for subsequent decisions and assessments, regardless of the validity of that anchor (Tversky & Kahneman, 1974). Thus, information provided by others, early impressions made, especially in situations with high mental load (Blankenship et al., 2008), can pose a problem regarding assessments, as ratings and assessments are undoubtedly processes which require great attention and exert high mental load and require high levels of vigilance from the rater's side (Adler, 2016).

These errors have a significant impact on a firm's operations (Kane et al., 1995). Often, performance ratings are coupled with pay systems in firms (Ileana Petrescu & Simmons, 2008). Thus, it is apparent that an inflation of or error in ratings, which are responsible for rewards, will eventually lead to a decoupling of performance and pay, not only making incentive systems less useful but also more costly if more people receive high ratings (Kane et al., 1995; Miceli et al., 1991). Furthermore, a dilution of the available data from ratings makes it more challenging to make accurate decisions when a firm is forced to downsize and to let employees go (Bernardin & Russel, 1993). Still, even more fundamental areas are affected by inaccurate performance reviews, like employee development (Mani, 2002). As values may serve as the basis for identifying the strengths and weaknesses of employees, inflated ratings can distort the image being portrayed and lead to the wrong allocation of resources, ultimately harming firm competitiveness and performance (Kampkötter, 2017; Kane et al., 1995; Trank et al., 2002). Also, if ratings are additionally the basis for promotions, making accurate

decisions once again becomes more crucial. This might lead to further dissatisfaction among employees, as the feedback they receive might not align with their organisational growth, including promotions or training and monetary, and non-monetary benefits they receive.

2.2.2 Tools and Interventions – Performance Assessments

Performance is being assessed all the time. Over the years several tools have emerged that tried to enhance accuracy of ratings and make them more reliable. One of the first tools used for assessing performance were order-of-merit scales that rank things based on their value, initially in the form of paired comparisons (Ortlieb et al., 2020). Later especially the military and public services collected evaluative data, which served as the foundation for Thorndike's (1920) classical work on errors in psychological ratings and Rugg's (1921) critique of the accuracy of ratings. This rising psychological interest in ratings had brought the field of ratings to flourish, and in the following years, several types of rating scales were constructed to counteract errors and issues.

Scales introduced in the following years include questionnaire scales, which present questions with either yes or no answer options or multiple answer opportunities (Raubenheimer, 1925). Weighted scoring scales were a further development of questionnaire scales. They paired questions/observations or points with an assigned score (Newcomb, 1929). Especially in psychological assessments, these weighted rating scales were often employed to assess and categorise illnesses, behaviour, etc. (Hartshorne et al., 1929; Terman, 1925). Frequency scales, which are a sub form of rated scales, follow the logic of normal distributions of social phenomena (Galton, 1869) and were first employed as a graphic scale by Symonds (1925). The scales mentioned above share a commonality, namely that they are graphic scales. These are scales that either describe a quality or pose a question, offering anchors and descriptions that aid the rater in making their choice (Weiss, 1933). Probably the most promising iteration on scales were behaviourally anchored scales, which provided descriptions of behaviours that coincide with scores and offered detailed insights into how a person is to be rated (Smith & Kendall, 1963). Unfortunately, a moratorium was called on the field of behaviourally anchored scales due to the unsatisfactory benefits in relation to the effort required to create them (DeNisi & Murphy, 2017).

Another popular intervention included rater trainings, which had a spike in popularity in the 1970s (DeNisi & Murphy, 2017). Initially training focused on lecturing raters, pointing out common issues that occur (Levine & Butler, 1952). However, pointing out what does not work, can lead to decreased

accuracy (Hedge & Kavanagh, 1988) and was therefore deemed as an ineffective approach (DeNisi & Murphy, 2017). The most widely cited training approach is FOR (Roch et al., 2011), which stands for frame-of-reference. This training involves viewing the appraised performance to be comprised of multiple dimensions, concretely defining the performance dimensions and creating prototypes for what the performance may look like (Roch et al., 2011; Woehr, 1994). Frame-of-reference training has been shown to increase accuracy through more accurate recall of the events and actions as well as creating a frame of reference, which makes approximating the true score more likely (Roch et al., 2011). True score in this sense refers to an established standard often set by experts (Woehr & Huffcutt, 1994).

Although scales and trainings try to remedy errors and even succeed to certain extent, appraisals still rely on the judgment of raters. Thus, in essence ratings are highly reliant on the perception and intentions of raters. Although trainings have shown more promising results than changes to scales, an important part of ratings has not been touched on, namely, informal appraisals like feedback. These trainings and scales also do not account for misaligned goals and incentives that may influence raters to give the ratings they do. Informal appraisals i.e. feedback and the political uses of appraisals will be explored in the next subsection.

2.2.3 Feedback and politics

Feedback is one of the key reasons for having appraisals (Cappelli & Conyon, 2017) and in theory, it is the main purpose of performance reviews (Carroll & Schneier, 1982). Feedback provides important information about performance and ideally compares set goals to the actual state of progress. This enables employees to adjust their behaviour and perform better subsequently. Though, in theory, feedback is something great and sometimes works in its intended manner, often it does not. Cleveland et al. (2007) found that providing as well as receiving feedback was a pain point for a lot of employees and was often met with disdain or at least discomfort (Adler et al., 2016). Unsurprisingly, feedback is received in a greatly different manner, depending on whether it is positive or negative, with positive feedback being better received and accepted (Landy & Farr, 1983). Negative feedback, on the other hand, can greatly affect employees and may even lead to insubordinate (Taylor et al., 1984) or even retributory (De Nisi et al., 1980) responses directed towards supervisors and peers, respectively. Furthermore, it is quite likely that people receiving feedback will face some sort of dissonance. The reason for this lies in the fact that people in general tend to rate themselves better in a variety of areas than the average (Alicke et al., 1995). This is not only true for more mundane aspects like driving

skills but also employee performance (Meyer et al., 1965). The above average effect in combination with the fundamental attribution error (Ross, 1977), which makes people internalise success and externalise failure, might further exacerbate feelings of unfair treatment. This is also supported by the findings of Pearce and Porter (1986), who found that people's attitude towards the organisation significantly worsened after receiving ratings labelled satisfactory, while a good review did not change the attitude of an employee towards the company. Further potential drawbacks of giving people bad or below-average reviews include loss of motivation, detachment, and self-isolation (Meyer, 1975). However, this effect may be mediated through companies' actively communicated efforts to improve appraisals (Landy et al., 1980).

How people react to feedback is also dependent on their type of orientation towards goals. Depending on a person's attitude, whether they are focused on learning or performance, the relationship to feedback is different (VandeWalle & Cummings, 1997). Learning orientation is characterised by wanting to build competence, acquire skills, and master them, whereas performance orientation is about demonstrating competence and acting in a manner that is perceived as favourable by others. People who are focused on learning and acquiring competence are more likely to exhibit feedback-seeking behaviour (VandeWalle & Cummings, 1997) and self-efficacy (Phillips & Gully, 1997), while people who fall under the category of performance orientation are negatively related to both. The way people react to feedback additionally depends on a person's behavioural plasticity, which describes the adaptability and ability to change behaviours based on outside stimuli (Brockner et al., 1988). The behavioural plasticity of a person also seems to be linked to self-esteem (Brockner et al., 1988) and self-efficacy (Gist & Mitchell, 1992) of an individual. The latter is also said to be related to performance (Gist & Mitchell, 1992).

Lastly, it must be mentioned that appraisals and feedback are not only used to find areas for improvement but also for political reasons (Murphy et al., 2004). Raters can follow several goals that include improving rapport with the rated person, using ratings as a signal for other ratees that one is being strict, hard to please with high expectations, or as a signal to other organisational units by exhibiting higher ratings and performance as a form of prestige (Murphy et al., 2004).

2.3 Meditation

Meditation has been practised in one form or another all over the world for millennia. Some Inuit tribes would carve a small circle repeatedly into a bigger stone to elicit a trance, tribes from the Kalahari Desert would vigorously dance in order to experience euphoria (West, 2016). Meditation is complex and can be done in a myriad of ways, shifting focus to singular sensory organs, focusing attention on objects or actions, reacting to and observing feelings or visualising and imagining constructs (Shear, 2006). The practices of meditation (concentration and mindfulness) that are most proliferated in today's Western world can be most closely linked to Hindu and Buddhist traditions (Sedlmeier et al., 2012). In the Western world, many forms of meditation have been secularised, as opposed to the way they are practised in Buddhism or Hinduism, which means they are not as strongly influenced by religious morals and beliefs (Kucinskas, 2014). Why meditation is done, excluding religious reasons such as enlightenment, can be broadly categorised into two reasons: First, tackling psychological and/or emotional issues and, second, enhancing one's understanding of their consciousness and embarking on a path of self-improvement and positive change (Coleman, 2002).

Often, these reasons are tied together and cannot be clearly distinguished, in part because they are often pursued in parallel (Coleman, 2002). Over the years, as meditation has become more established and researched in the Western world, it has been more frequently used as a psychotherapeutic tool (Walsh & Shapiro, 2006). It has been mentioned that it was used as a clinical tool in North America as well as Europe as early as 1979 (West, 1979). The reason meditation is so common is its numerous positive health benefits. It has been shown to reduce stress (Ospina, 2008), anxiety (McGee, 2008), addiction (Alexender et al., 1994), depression (Kabat-Zinn et al., 1992), chronic pain (Kabat-Zinn, 1982) and insomnia (Gong et al., 2016). Meditation has been shown to have not only positive psychological but also physiological effects on the body. Studies have shown meditation can lead to decreases in heart rate (Cahn & Polich 2006; West, 1978), respiration rate and oxygen consumption (Sugi and Akutsu, 1968), and decrease in skin conductance (Woolfolk, 1975). In general, this aligns with findings of lowered activity in the sympathetic nervous system and heightened activity in the parasympathetic system (Benson et al., 1974; West, 1979). The sympathetic nervous system regulates the body's fight-or-flight response, which is triggered by stressful situations, while the parasympathetic system is responsible for activities related to rest and digestion (Jänig, 2003).

Bluntly, meditation seems to relax the body. Western literature has so far categorised forms of meditation based on three criteria: attention, cognition, and the aim of the practice (Walsh & Shapiro,

2006). According to Walsh & Shapiro (2006) the criteria interplay as follows. Attention allows for a differentiation to be made between concentration and awareness. While concentration refers to one-pointedness of attention either on a singular object or one's breath, awareness is more fluid and allows for the focus to wander between objects consecutively. Cognition refers to the relationship of the meditator to his thought processes. While some practices focus on observing thoughts as they emerge, others actively try to shape them. Lastly, meditation practices differ in their goals. Some are more holistic and aim to enhance overall well-being. In contrast, others are more specialised and are meant to improve qualities, such as compassion (loving-kindness) and concentration (forms of Zen meditation). This also means that different types of meditations have different effects on their practitioners and can potentially lead to differing results or the development of qualities at a different pace (Walsh & Shapiro, 2006).

For the sake of completeness, it needs to be stated that meditation has also been shown to have adverse effects, though the downsides of meditation have been less frequently documented in the literature. Still, a project called *Varieties of Contemplative Experience* by researchers from Brown University has been concerned with uncovering the negative aspects of meditation, resulting in 19 publications to date. Some of these effects include worsening of adverse effect such as anxiety and other anomalies (Farias, 2020). Other more severe cases include changes in the self and awareness that lead to serious impairments, like depersonalization and disassociation (Lindahl & Britton, 2019).

In the context of this thesis the focus lies on a form of meditation, namely, loving-kindness, which will be further detailed in the next section.

2.4 Loving-kindness Meditation

Loving-kindness is a form of Buddhist meditation and aims to foster benevolence and acceptance (Harvey, 2012). Primarily, it focuses on helping users develop good intentions and qualities, such as patience, openness, and friendliness, while simultaneously, emphasising prevention of evil or ill thoughts that negatively affect one's mind (Harvey, 2012). This meditation is practised in a manner that involves chants, phrases and visualisation (Reilly & Stuyvenberg, 2023). First, the meditation brings attention to the self. Then, the phrases and chants reinforce the notion of self-acceptance, including all one's faults, and self-love. Afterwards, the focus shifts to the outside realm and love, acceptance or very commonly blessings/benevolent wishes are directed towards others. Phrases and chants that are used could be as follows: "May I be well and happy, may I be free from difficulties and troubles" and that somebody else may be happy and free from troubles as well (FitMind, 2021).

Loving-kindness has been shown to increase compassion (Kearney et al., 2013; Kirby & Laczko, 2017). Compassion involves not only caring and concern but also the intent to alleviate suffering (Goetz & Simon-Thomas, 2017). This would make it distinct from kindness, which puts a focal point on creating happiness (Gilbert et al., 2019). Other definitions also include self-kindness and a non-judgmental attitude towards oneself (Neff, 2003; Germer, 2009). Regardless, both compassion and kindness, though slightly dissimilar, promote individuals' well-being (Gilbert et al., 2019). Loving-kindness has been linked to the improvement of mood and the fostering of positive emotions (Galante et al., 2014). It has also been shown to continuously increase mood by providing a constant influx of positive emotions (Diener et al., 2006). Although it is worth noting that meta-analyses have found significant variation in effect sizes, the effects are generally positive (Zeng et al., 2015; Reilly & Stuyvenberg, 2023). However, this is partly understandable as meditation research is inconsistent regarding the duration, type, and controls for interventions, with most interventions usually being longer (Zeng et al., 2015). It is also worth noting that loving-kindness meditation has been shown to have a greater impact with larger effect sizes concerning positive mood (Hafenbrack et al., 2020) than for example mindfulness meditation.

Furthermore, research conducted so far has utilised guided meditations, similar to the study presented in this thesis. This thesis' study involved a five-minute-long guided loving-kindness meditation, after which people's affect was measured. Based on the audio and the focus of the meditation, whether it is compassion, happiness, or kindness experienced effects may differ (May et al., 2016). Loving-kindness has far-reaching effects, and the technique has been shown to reduce the severity of

depression and enhance life satisfaction and personal well-being by generating positive emotions (Gu et al., 2022; Lv et al., 2024). This thesis postulates that these positive effects are measurable despite the brevity of the guided meditation. Therefore, the following hypothesis was postulated:

H1. Loving-kindness meditation increases positive affect compared to the control condition.

Loving-kindness has been further shown to decrease several negative interpersonal attitudes, such as anxiety, biases, judgments, anger and aggression (Zhou et al., 2023). Therefore, this thesis also postulates that:

H2. Loving-kindness meditation decreases negative affect compared to the control condition.

2.5 Mindfulness

Mindfulness stems from the Pali word sati, which translates to remembering or lucidity of mind (Davids & Stede, 1959). It describes a state of mind characterised by wakefulness and awareness. Awareness of one's surroundings and one's internal thought processes and emotions enables one to create distance between oneself and arising thoughts (Siegel et al., 2009) and (re)direct attention to regulate one's feelings (Singh et al., 2003). At the same time, mindfulness not only creates distance but allows one to stay curious and open towards arising thoughts (Bishop et al., 2004). Mindfulness has its roots in Buddhism and is an integral part of Buddhist meditation (Kabat-Zinn, 2003). However, Kabat-Zinn (2005) also notes that meditation can be regarded as a tool to foster mindfulness. Mindfulness is considered the fundamental function which allows one to recognise which qualities to cultivate in oneself (Rapgay & Bystrisky, 2009).

In the Western world, mindfulness gained popularity largely due to Kabat-Zinn's clinical work, specifically MBSR, which stands for Mindfulness-Based Stress Reduction (Kabat-Zinn, 1990). The impact of this work is visible in the number of citations, which, as of this date, numbers over 19,500, as shown on Google Scholar. MBSR is an intervention used in clinical psychology to treat psychological and physical symptoms like anxiety, distress or pain (Kabat-Zinn, 1990; Baer, 2003). There are several theories on how mindfulness works. For example, Shapiro et al. (2006) propose that mindfulness works through intention, attention, and attitude, as well as an additional aspect, which is on the meta level called reperceiving. Intention refers to a vision of where one aims to go with the help of mindfulness. These intentions can also shift with increased experience (Sedlmeier et al.,

2012). Shapiro (1992) found that people's aims/intentions shift from regulating attention to exploring inner workings and lastly to not being bound by a definition of self. Attention refers to observing both the outside world and internal processes in every moment. An important distinction must be made that it is not about interpreting but observing and living in the moment (Shapiro et al., 2006). Attitude refers to the way one approaches awareness. This could be either an analytic approach or an approach characterised by compassion and acceptance (Shapiro et al., 2006). However, the latter approach is more advisable and serves as the foundation of mindfulness (Kabat-Zinn, 2003; Shapiro & Schwartz, 1999). Lastly, the meta-mechanism of reperceiving is tied to a shift in perception from subject to object, which translates into a more neutral perception of one's circumstances and inner workings (Shapiro et al., 2006). This is also mentioned in Kegan's (1982) work *The Evolving Self* to be a key element for developing the ability to construct meaning and to understand the world by detaching oneself from oneself and seeing the world from the viewpoint of others.

Later approaches to conceptualise key aspects of mindfulness systematically include highly similar items. For example, the four factors proposed by Feldman et al. (2007). These focus on (1) directing attention to the (2) present moment, combined with (3) awareness and a (4) non-judgmental, open-hearted attitude (Feldman et al., 2007). There are also different types of mindfulness, either trait, which describes a natural inclination towards mindful behaviour in everyday life (Baer et al., 2006), or state mindfulness, which is a state of heightened awareness and mindfulness often, but not solely, induced by meditation (Lau et al., 2006). Depending on the type of mindfulness, different conceptual frameworks, measures, and tools may be more useful. For example, for trait mindfulness MAAS, The Mindful Attention Awareness Scale, assesses states of mindfulness over time with a 15-item questionnaire (Brown & Ryan, 2003). On the other hand, state mindfulness is assessed with the help of the TMS, Toronto Mindfulness Scale, which is centred around two main dimensions, decentering and curiosity, and has been shown to be an adequate measure and predictive tool for treatment outcomes (Lau et al., 2006). In current literature, most meditation treatments to foster mindfulness concerned long-term interventions with varying length up to a few weeks (Zeng et al., 2015). However, it has been shown that even shorter forms of meditation can significantly increase state mindfulness dimensions (see Hafenbrack et al., 2022; Zhou et al., 2023). Still, even then most meditations lasted around 15 minutes or longer which may not be practicable in a work setting. Therefore, the study conducted in the context of this thesis involved a five-minute-long meditation. Thus, this thesis hypothesises that:

H3. Loving-kindness meditation increases mindfulness dimensions, curiosity and decentering, compared to the control condition.

Mindfulness can also serve as a valuable skill that benefits individuals in various areas of life. Mindfulness has been shown to reduce multiple biases. Maymin & Langer (2021) have demonstrated that mindfulness can reduce several cognitive biases, including availability heuristics, overconfidence, anchoring, and confirmation bias, which, in the context of performance appraisals, can potentially play a role. Thus, this thesis postulates that:

H4. Loving-kindness meditation, through the indirect effect of increased mindfulness and its components, curiosity and decentering, reduces the halo effect in performance appraisals.

To test this additionally to measuring affect participants were presented with a video and one of two possible descriptions of a person which served as a basis for a performance appraisal.

3 Methodology

The experiment conducted in this thesis aims to shed light on the effect of state mindfulness on mitigating rater biases, specifically the halo effect. The experiment thus aimed to assess the impact of state mindfulness, characterised by a present focus, empathy, a nonjudgmental attitude, as well as a positive mood elicited by loving-kindness meditation, on rater validity.

3.1 Participants – Recruitment and Demographics

The experiment was conducted with Qualtrics. Participants volunteered and participated anonymously. Most people were recruited through personal contacts, university groups, and cooperation with student associations. These participants received no monetary incentives for their participation. To meet an adequate sample size of at least 130 participants for analysis (Brysbaert, 2019) further 40 participants had to be recruited through Prolific. These latter participants were offered a monetary incentive of 2.30 €. Participants recruited through Prolific's standard sample were screened for English proficiency by Prolific. An additional field to capture people's Prolific ID was added to the experiment and the amount of payment was set in accordance with Prolific's guidelines.

A total of 255 responses were collected. However, a large portion of these responses did not provide complete data. Out of the 255 responses, 170 participants completed the experiment, and 85 participants aborted it. It is worth noting most of those who have not completed the survey have left Qualtrics in under a minute and have not progressed beyond the first attention check, indicating they have not even been presented with the meditation or control audio. Demographic data was only available for 24 out of the 84 non-completed responses, which were predominantly from female respondents (70.83%) and students (58.33%).

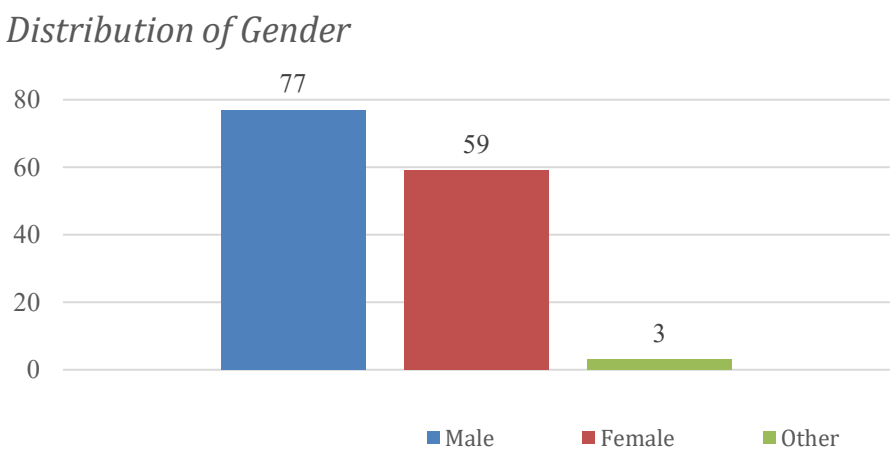
Furthermore, responses were excluded when any value was missing in the performance items, resulting in the removal of eight additional responses. An in-depth look at the data revealed that several participants took longer than was explainable through technical difficulties, meaning that they either had taken a break during filling out, set the device aside or simply did not pay attention. Therefore, participants who exceeded double the intended maximum fill-out time of 40 minutes were excluded from the study. Thus, 14 additional responses were removed. Lastly, the two attention checks were assessed. Entries with incorrect answers to the attention checks were further filtered and

excluded from analyses. From the 148 responses, only nine failed the attention checks, resulting in a final number of 139 valid responses for analysis.

From the 139 participants who completed the survey, 74 were male, 60 were female, and three identified as other (See Figure 1).

Figure 1

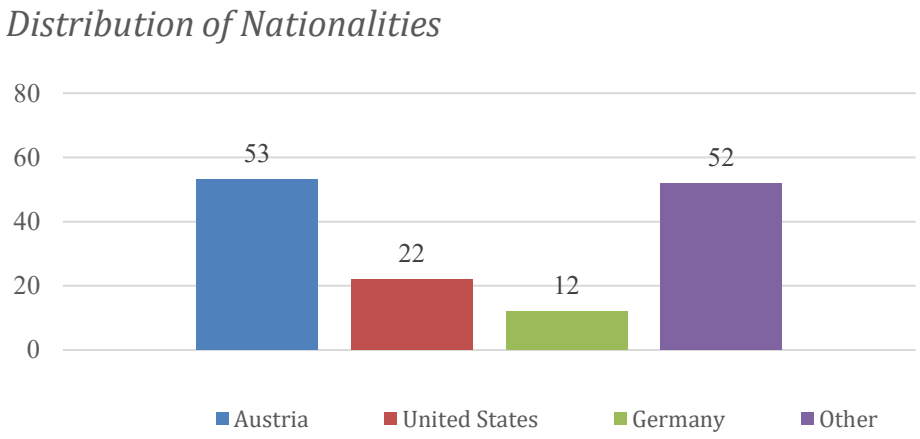
Distribution of Gender



In total, people from 36 nationalities completed the survey, with the majority being from Austria, the United States, and Germany, accounting for 62.59% of all participants. Austria was the largest group, comprising 38.12% (See Figure 2). The other 33 nationalities are combined into the subgroup "Other".

Figure 2

Distribution of Nationalities



Most participants were students (46.76%) or employed in a full-time position (31.65%) (See Figure 3). Smaller subgroups are displayed as "Other" - comprising, listed in descending order, part-time employees, retirees, and individuals seeking employment. Participants were aged between 18 and 77 with the overwhelming majority being aged below 34 (79.86%) (See Figure 4).

Figure 3

Distribution of Occupation

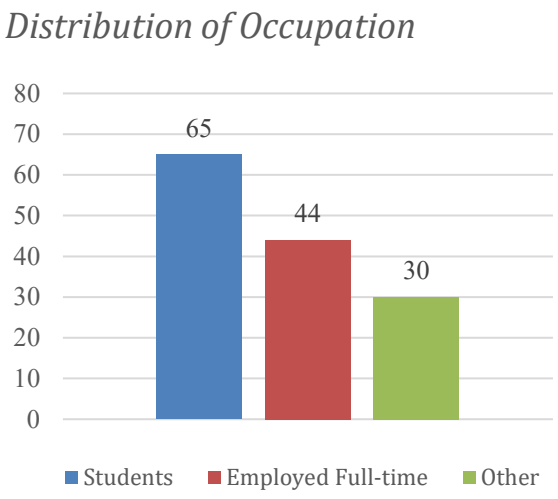
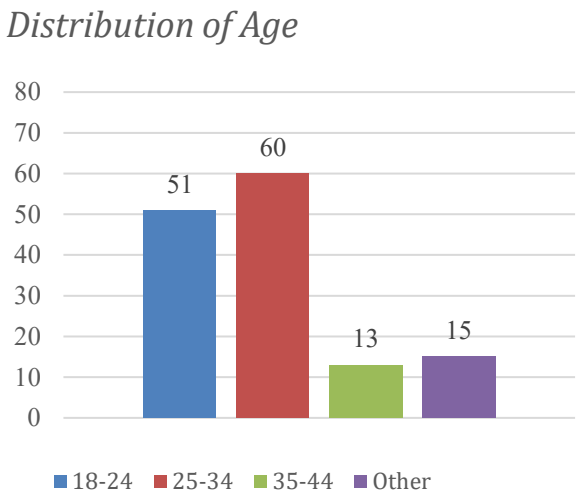


Figure 4

Distribution of Age



Due to the removal of invalid and non-finished responses, there is an imbalance regarding the randomization (See Table 1). This means that subsequent analyses especially on the level of the moderated mediation, must be interpreted with caution.

Table 1

Randomization Table - Induction and Description type

Group	Number	Percentage of total participants
Experimental Condition	63	45.32%
Control Condition	76	54.68%
Experimental Condition x Neutral	35	25.18%
Experimental Condition x Moral	28	20.14%
Control x Neutral	37	26.62%
Control x Moral	39	28.06%

3.2 Experimental Setup

Participants were first presented with information about the estimated duration of the experiment and that their participation was voluntary. Furthermore, a disclaimer was placed on the first page that the form should be filled out in a quiet environment, ideally with headphones. Initially, the experiment's setup required participants to complete a demographic questionnaire on the last page of the study. This included information regarding age, educational level, nationality, gender, ethnicity, occupation and income. Due to high dropout rates, the demographic questionnaire was placed at the beginning of the survey to allow for a deeper examination of the demographics of individuals who do not complete it. A short audio clip, which played the word “Apple”, was placed at the beginning of the experiment to allow participants to check their audio and serve as an attention check.

Participants were then randomly assigned to either the control or the experimental condition. Participants in the experimental group were presented with a five-minute-long loving-kindness induction. A similar recording, albeit longer, was used in Hafenbrack et al. (2020), Hutcherson et al. (2008) and Salzberg and Kabat-Zinn (2004). The audio recording is a guided meditation by Sharon Salzberg, a leading expert in the field of loving-kindness meditation (see link in the Appendix 8.1). Participants in the control condition were presented with an excerpt from a NY-Times article in audio form of the same length as the meditation. The article discusses the 2014 rethinking of the SAT exam (Lewin, 2014). This particular recording has been employed before by Hafenbrack et al. (2020). Participants were not able to proceed with the form until the audio had finished. After the audio was completed, all participants were presented with a further attention check, which contained four potential options, which were chosen to be similar to each other but still distinct enough to avoid confusion among participants:

- The audio focused on practising slow and steady breathing.
- The audio discussed proposed changes to college classes.
- The audio emphasised spreading kindness and love.
- The audio talked about making adjustments to the SAT exam.

Participants were then presented with the Toronto Mindfulness Scale (TMS; Lau et al., 2006). The TMS enables the testing of the effectiveness of the mindfulness induction, measuring two sub-scales, decentering and curiosity/openness, on a five-point Likert scale. It contains 13 items, of which six measure curiosity/openness and seven measure decentering. It has already been used in multiple

studies as a measure of state mindfulness (see: Bunjak et al., 2022; Erisman & Roemer, 2010; Hafenbrack et al., 2020; Lau et al., 2006).

The participants were then asked how they felt in the current moment. The time frame for the questioning, i.e., the present moment, was chosen to assess the effect of loving-kindness meditation on affect. The scale employed to measure participants' affect was the PANAS-SF (Positive and Negative Affect Scale - Short Form; Thompson, 2007), a more focused and concise version of the PANAS (Watson et al., 1988). Among other items, it contains the following: Interested, Distressed, Excited, Upset. In total, it includes 20 items, of which ten are indicative of positive emotional states and another ten are indicative of negative emotional states, measured on a five-point Likert scale.

Both scales, PANAS-SF and TMS, were labelled and ranged from "not at all" to "moderately*" to* "very much". The two measures were presented in a random order to avoid any potential effect that might arise from the order of questions.

Afterwards, participants were presented with a short description of a fictitious professor. Participants were randomly assigned to one of two optional conditions for the task: a neutral and a moral condition. Both conditions presented the participant with a description of the teacher, which included the teacher's age, education, a photograph (as visible in Figure 5), and a brief text.

Figure 5

Picture of Experimental Subject

The text included the following passage for all participants: "Your professor is a 28-year-old male who is currently writing his Ph.D. at the University of Wisconsin and was born in Columbus, Ohio, USA*".* The text was manipulated to add further information in the moral condition, which supported the fictitious professor's character. For this portion of the task, the same manipulation was applied as

in the third study of Stellar and Willer (2018), with two changes to the original method. First, the neutral condition included sentences designed to be similar to the moral conditions, naturally without a moral component, to exclude any differences in study duration between conditions. Second, the sentences were merged into a text and presented simultaneously in a manner more similar to a professor's biography, rather than being presented as a list with bullet points. An example of a sentence from the moral condition is “Furthermore, the student magazine of his university reports that he donated a large amount of money to a charity for homeless kids”. In contrast, the neutral condition of the same sentence was “Furthermore, the student magazine of his university reports that he usually goes for walks around the university campus after lunch”.

After participants received information about the professor, they were presented with a short video purportedly created by the professor. Participants were presented with the same video regardless of previous conditions or treatment groups. The task of the participants consisted of rating the professor’s teaching performance based on the short video. The video provided explanations of game-theoretical concepts and was selected based on the likely unfamiliarity of participants with the topic and the relative ease of comprehensibility and digestibility of the subject matter. Furthermore, it is worth noting that the choice of video presented a challenge, as the tone of voice, audibility, view count, production value, and numerous other aspects of the video had to be considered. This is because a video that is already very well-known and well perceived by audiences who have already watched it would likely skew ratings due to its high quality. Thus, the video recording had to be minimalist, containing only a voiceover and slides, and not overly polished. Similarly to the meditation audio recording, participants were also unable to proceed until the video was completed. After viewing the video, the professor was then rated based on a slightly adapted version of Harari and Zedeck’s (1973) scale for evaluating faculty teaching. The scale used to assess performance consists of five items, each measured on a 7-point Likert scale ranging from “very poor” to “excellent”. Performance was measured along the following dimensions: delivery, depth of knowledge, structure of the lecture, relationships with students, and fairness & grading. Participants were provided with clarifications as to what each dimension represents and includes. For example, fairness & grading refer to the transparency of grades and the thoroughness and timeliness of feedback. Descriptions like this were provided for all dimensions. All experimental materials, including the survey and links to the files used can be found in the Appendix.

4 Results

Analysis was conducted using the statistical program R and R Studio, with the following packages: tidyverse, easystats, gtsummary, readxl, ggplot2, psych, corrplot, corrr, hmisc, and dplyr. The data were first cleaned by removing responses that did not progress to the ending page. Additionally, responses that contained missing values (NA) in any of the scales or measures, responses where the participant failed at least one of the two attention checks, and durations, which exceeded double the intended maximum fill-out time (40 mins) were removed. Additionally, people whose time between the first click and last click was of a shorter duration than the length of the audio recording were removed, as this indicated that they had not listened to the audio which requires to be clicked to be started. Frequency tables were calculated for completed and incomplete submissions, failed attention checks (from completed submissions) and demographic data. Furthermore, demographic data were visualised to display the distributions of age and gender for the treatment and control groups, to assess whether randomisation within groups was adequate.

4.1 Preliminary Analysis

Cronbach's α was calculated for all scales, including TMS, PANAS-SF, and the performance measures, to determine the reliability of the items. Internal consistency is assumed to be acceptably ensured if Cronbach's α is equal to or above .70. If this holds, the scores of the scales can be used for further analyses, in line with research by Bland & Altman (1997). As Cronbach's α is a unidimensional measure, the PANAS-SF and the TMS had to be analysed on a subscale level. At the same time, performance measures were analysed as a single scale. Cronbach's α was above .70 for all (sub)scales (see Table 2 for results), ensuring internal validity of measures. Especially curiosity and the affect scales had excellent validity.

Table 2

Cronbach's α for employed scales and subscales

Scale/Subscale	Items	Cronbach's α
Performance Measures	5	.80
TMS	13	
Decentering	7	.72
Curiosity	6	.91
PANAS-SF	20	
Positive Affect	10	.90
Negative Affect	10	.87

Additionally, to Cronbach's α, a factorial analysis was conducted to measure the distinctiveness of items across all scales. In order to test the data's suitability for factorial analysis, a Kaiser-Meyer-Olkin (KMO) index was calculated, and a Bartlett's Test of Sphericity was conducted. The KMO and the sphericity tests are prerequisites for conducting factorial analysis (Kaiser, 1974; Bartlett, 1950). The KMO measures common variance in variables, it is a comparison of correlations and partial correlations between items. If KMO is above .60, variables are deemed suitable for further analysis (Kaiser, 1974). Bartlett's Test of Sphericity, on the other hand, compares a correlation matrix of the variables to the identity matrix, checking for significant differences. If the difference is substantial, with $p < .05$, which means that items are correlated to each other, factor analysis is appropriate (Bartlett, 1950). Both tests result in a value between 0 and 1 with higher values indicating suitability for factor analysis. KMO was consistently above .60, and Bartlett's was significant for each scale as well as subscale with p-values < .001. A complete list of all values corresponding to the analyses is presented in Table 3.

Table 3

Preliminary analysis - KMO and Bartlett's Test of Sphericity

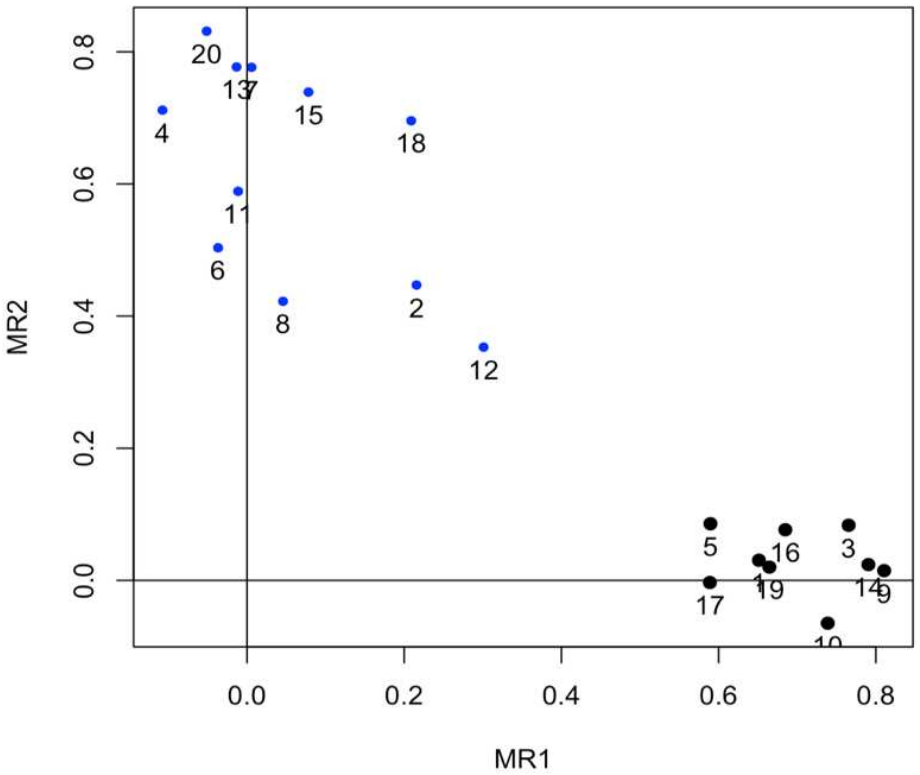
Scale/Subscale	KMO MSA - scale	KMO MSA- subscale	χ^2 (Bartlett)	*df*	*p*
Performance Measures	.77	-	212.04	10	< .001
TMS	.89	-	769.89	78	< .001
Decentering	-	.73	171.76	21	< .001
Curiosity	-	.90	497.13	15	< .001
PANAS-SF	.86	-	1376.31	190	< .001
Positive Affect	-	.90	708.43	45	< .001
Negative Affect	-	.88	558.58	45	< .001

All scales were, thus, deemed suitable for further analysis, and most of them can be categorised as excellent, judging by the MSA, measure of sampling adequacy, for the KMO. Bartlett's showed that, especially for the PANAS-SF, factor analysis was suited and deemed valid for all other scales as well. The subscales exhibited strong intercorrelation, further justifying the conduct of factor analysis.

Subsequent factor analysis validated the structure of PANAS-SF. Figure 6 represents the factor loadings of the PANAS items for the identified factors. Two factors were identified, these are labelled MR1 and MR2 in Figure 6. The values on the y and x axis of the graph indicate the loadings of each variable to the corresponding factor. A full list of loadings and variables can be found in the Appendix 8.3. Some individual items revealed cross-loading, meaning that they were attributed to both categories. Most variables were assigned to the accurate subscales. In Figure 6 on the upper left side, in blue, MR2, the negative affect scale, and on the bottom right, MR1, in black, the positive affect scale variables are visible. The item "Alert" (12) exhibited the most ambiguous position, which could be due to the reason that as though it is a positive affect item, alertness may be associated with a pertinent danger and a feeling of being on edge, which may lead to confusion in interpretation by participants. Regardless, it loaded more onto the factor which coincides with positive affect. Thus, the two dimensions of the PANAS-SF were discernible.

Figure 6

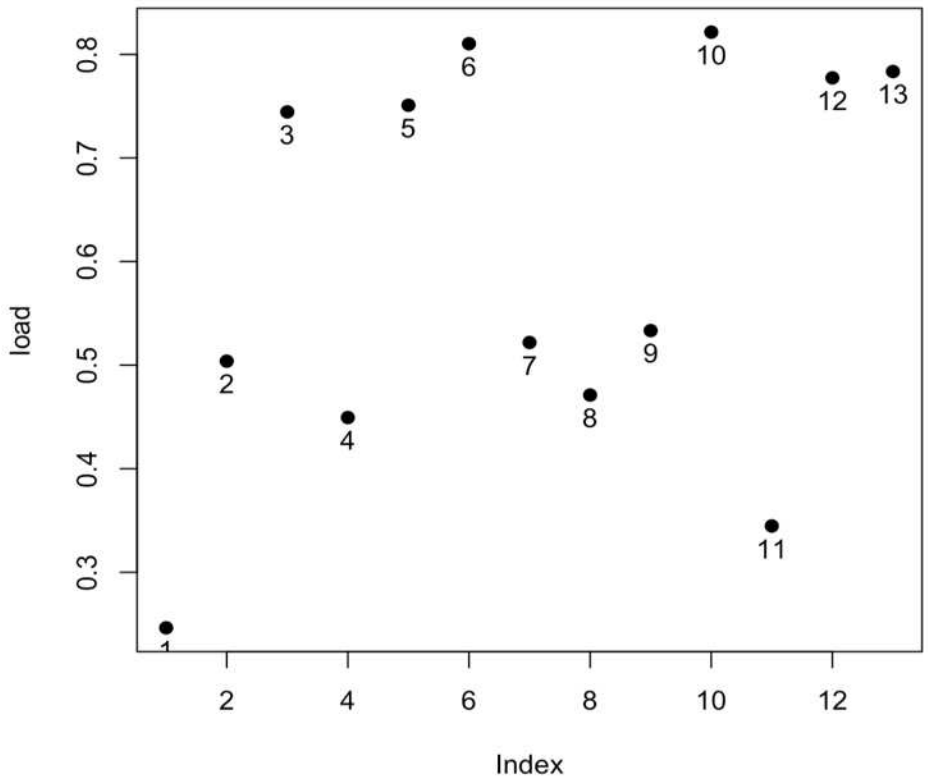
Factor Analysis Clustering - PANAS-SF Items



However, no clear distinction was found between TMS subscales (see Figure 7). They were loaded to the same factor. On the y-axis the strength of the loading to the identified factor can be seen, ranging from 0 to 1. The x-axis is simply an index of the variables and is of no significance for interpreting the graph. However, the reason for only one factor being identified may be because opposed to the PANAS-SF, which measures orthogonal scales, the TMS measures parts of an overarching effect, namely mindfulness. However, the items exhibited different levels of loading, within the single identified factor. The variables that have higher levels of loading with the identified factor belong to the dimension curiosity whereas the variables which exhibited lower levels of loading with the identified factor belong dimension decentering. Regardless, only one factor has been identified for this scale.

Figure 7

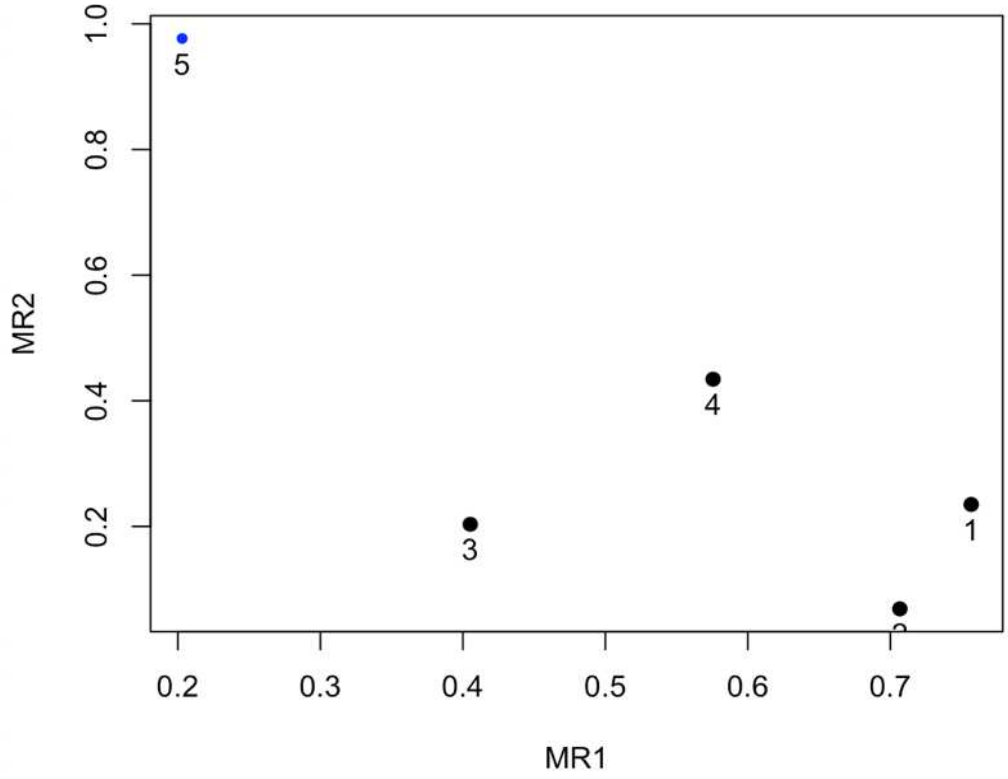
Factor Analysis Clustering - TMS Items



Analysis of the performance measures revealed two factors. Although there was cross-loading between most items, the performance measure 5 - “Grading” is identified as its own factor. This indicates that it measures a different aspect compared to the other variables. The experimental design only allowed direct assessment of three of the rated criteria, namely, delivery (1), knowledge (2) and structure (3), but did not provide any information on the dimensions relations (4) and grading (5). Thus, ratings for the last two variables had to be inferred by the participant and were more likely to be influenced by the mental image they had formed of the person. Thus, this could serve as an explanation as to why grading was identified as its own factor.

Figure 8

Factor Analysis Clustering - Performance Measures



As Cronbach’s α was sufficiently high, supporting consistency of the scale, and factorial analysis sufficiently supported the validity of the underlying constructs measured, individual items were transformed into scale scores by summing the items of the scale. Subsequently, a bivariate analysis was conducted to examine the relationships between demographic variables and continuous variables. This was also done to check for any systematic differences between genders or nationalities in responses. To conduct the analysis, gender, occupation, and nationality were transformed into dummy variables which represent the two most frequent response types in each variable. When a response category accounted for approximately 50% or more of the responses, it was maintained as a single category.

Table 4

Bivariate Analysis Results

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. Induction	-														
2. Prolific	-.02	-													
3. Female	.08	-.16	-												
4. Student	-.10	**-.34**	-.07	-											
5. Austrian	.00	**-.39**	.18	**.30**	-										
6. Age	.11	**.38**	-.16	**-.42**	**-.32**	-									
7. Decentering	.17	-.09	-.03	.04	.03	.01	-								
8. Curiosity	**.30**	.01	.10	-.04	.02	.01	**.61**	-							
9. Pos Affect	**.22**	.01	.15	-.09	.06	.03	**.40**	**.45**	-						
10. Neg Affect	.04	-.07	.20	.04	**.24**	-.19	-.04	.02	.05	-					
11. Delivery	-.10	.19	-.03	-.13	-.09	.08	-.02	.03	.05	-.07	-				
12. Knowledge	-.08	.21	-.09	-.13	.03	.02	-.03	.07	.11	**-.22**	**.50**	-			
13. Structure	.00	.06	.09	-.03	.06	.12	.05	.09	.10	-.15	**.48**	**.42**	-		
14. Relations	.06	.05	**.22**	-.11	-.02	.08	.08	.15	.15	-.04	**.55**	**.43**	**.42**	-	
15. Grading	.17	.10	.18	-.06	-.01	.15	.16	.21	**.27**	-.13	**.40**	**.24**	**.44**	**.59**	-
16. Halo Score	-.11	-.01	-.10	-.08	.08	-.17	-.15	-.10	-.20	.04	-.18	.11	-.15	**-.50**	**-.59**

*Formatting indicates significance: $p < .05$ = 0, $p < .01$ = **0**, $p < .001$ = **0.**

Bivariate analysis (see Table 4), in the survey context revealed that people from Austria were more likely to be students compared to participants from other countries. An additional finding was that negative affect significantly and positively correlates with the dimensions Austrian, Female and negatively correlates with Age. This means that Austrian people and females on average reported higher negative affect than people from other countries. On the other hand, younger people reported lower negative affect. Although prolific users are older, have a different occupation, and nationality, they displayed no significant differences in positive or negative affect, but rated the person slightly higher on Delivery and Knowledge. Females also tended to rate higher on the Relations dimension. Performance measures (11-15) were all positively correlated with each other. Decentering correlated positively with Curiosity and both variables positively correlated with Pos Affect. Decentering and Pos Affect were negatively correlated with halo score. Curiously, at the same time Decentering, Curiosity and Pos Affect all positively correlated with Grading. Relations, Grading and Delivery also

negatively correlated with Halo Score. Neg Affect significantly negatively correlated with Knowledge.

A further step in the analysis involved assessing the skew and distribution of the collected continuous variables related to affect and state mindfulness. The x-axes of the Figures 9-12 are scaled to show the minimum and maximum values possible on the respective scale. Curiosity (see Figure 9) is relatively uniformly distributed with smaller peaks in the mid-range. Decentering (see Figure 10) scores are slightly skewed, though it must be noted that observations do not range to the maximum of this subscale. This might be due to the items of decentering being harder to grasp than curiosity. Decentering involves more abstract language (e.g., "I experienced my thoughts more as events in my mind than as a necessarily accurate reflection of the way things 'really' are"). Participants in this study have given feedback about sometimes not understanding descriptions of individual items belonging to the decentering subscale due to their abstractness.

Figure 9

Histogram - Variable Curiosity

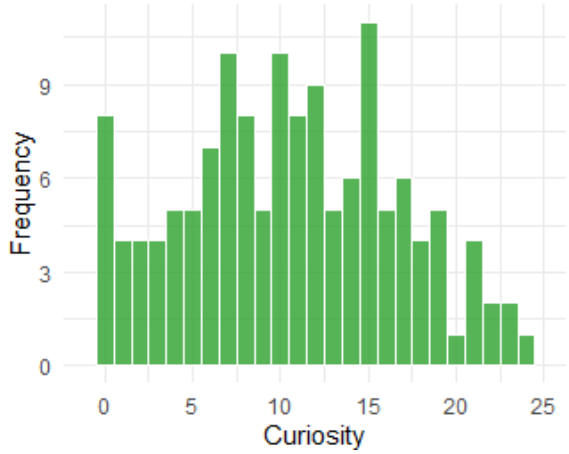
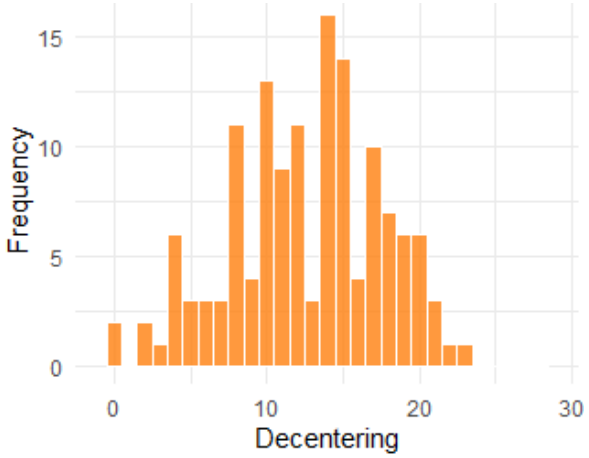


Figure 10

Histogram - Variable Decentering



Whereas positive affect (see Figure 11) is relatively symmetrical, the distribution of negative affect (see Figure 12) is greatly left-skewed, meaning that most participants experienced very low levels of negative affect.

Figure 11

Histogram - Variable Positive Affect

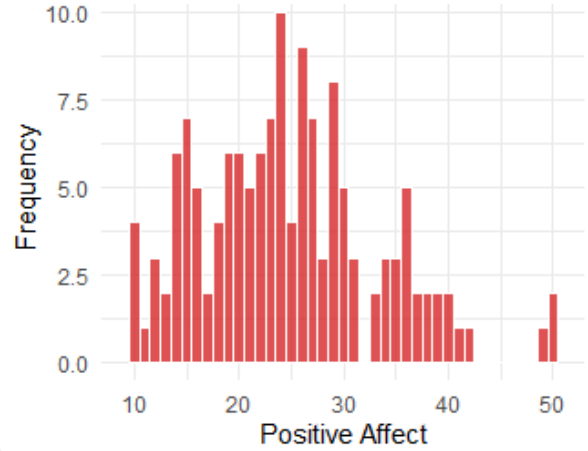
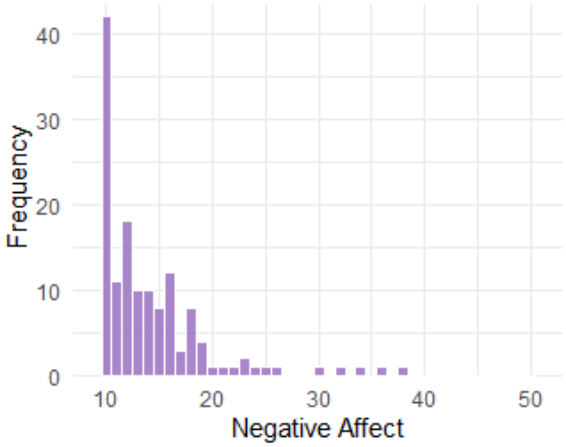


Figure 12

Histogram - Variable Negative Affect



4.2 Hypothesis Testing

H1 postulated that loving-kindness meditation increases positive affect compared to the control condition. Tied to this, H2 postulated that loving-kindness meditation similarly would decrease negative affect compared to the control condition. This was also analysed with a t-test (see Table 5). Meditation significantly increased self-reported positive affect. Therefore, H1 was supported. For negative affect, no significant effect was detected, so H2 was not supported.

Table 5

Results - Independent t-tests for affect variables depending on induction

	Induction		Control		*t*	*df*	*p*
Variables	***M***	***SD***	***M***	***SD***			
Positive Affect	26.76	8.56	23.01	8.23	2.61	130.15	<.01**
Negative Affect	14.38	5.54	13.92	5.11	0.50	127.71	.62

*Stars indicate significance: $p < .05$ = *, $p < .01$ = **, $p < .001$ = ***

H3, which measures the effectiveness of the induction, was tested with a t-test (see Table 6). Curiosity was significantly higher in the induction group with a p-value < .001. Decentering was also significantly higher in the induction group, at a p-value of .045. Thus, the mindfulness induction was successful and H3 was supported.

Table 6

Induction Test - Independent t-test

	Induction		**Control**		***t***	***df***	***p***
Variables	***M***	***SD***	***M***	***SD***			
Decentering	13.46	4.27	11.80	5.40	2.01	136.72	.045*
Curiosity	12.53	5.55	8.90	6.01	3.70	135.38	<.001***

*Stars indicate significance: $p < .05$ = *, $p < .01$ = **, $p < .001$ = ***

A preliminary analysis for H4 was done by examining intercorrelation heatmaps between groups. The heatmaps (see Figure 13 and 14) show Spearman's rank correlation coefficient for each variable-pair of the performance dimension by treatment group. The coefficient ranges from -1 to 1 indicating the direction and strength of the correlation. The heatmaps below show only positive coefficients meaning that the relationship between the variables is positive. Thus, an increase in one coincides with an increase in the other.

Still, a reduction of intercorrelation of specific items of around 0.1 could be detected. Especially between items that are in essence less related dimensions, such as between grading and all the other items, as has been ascertained through factor analysis. The intercorrelation in the treatment group only increases between knowledge and relations to students. The average intercorrelation in the control group is at .49 whereas in the treatment group it is at .41. Most significant reductions were measured in the intercorrelation of grading with other items which is at .49 in the control group and at .35 in the treatment group.

Figure 13

Correlation Matrix - Heat Map control group

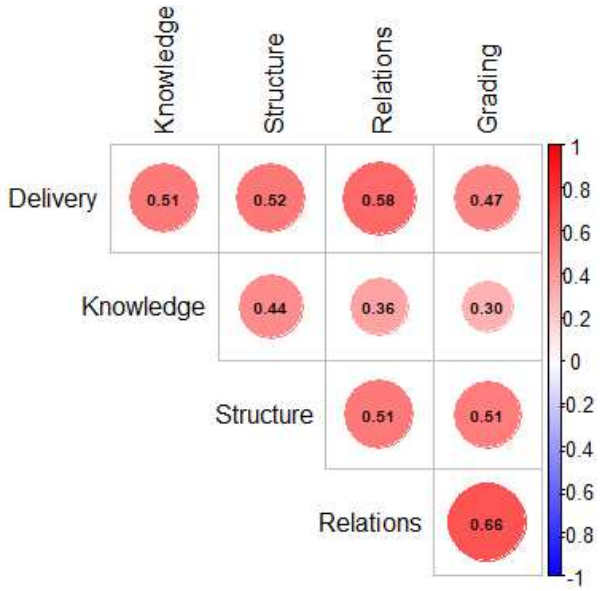
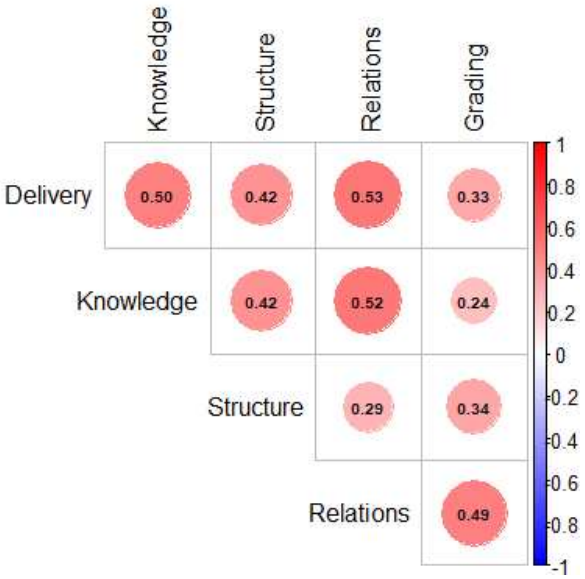


Figure 14

Correlation Matrix - Heat Map treatment group



However, the preliminary analysis does not allow inferences about significance due to the significant drop in observations, making t-tests, ANOVAs or Wilcoxon tests ineffective. This challenge in assessing halo highlights an issue with evaluating halo bias, specifically since the intercorrelation of items cannot be calculated on an individual level. This problem has also been discussed in Fisicaro & Lance (1990). Thus, a new variable had to be created to operationalise halo on an observational basis. This was done using the method applied by Palmer & Loveland (2008) and Balzer & Sulsky (1992). This method operationalises halo by the individual squared discrepancy to the mean of an observation.

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Where n stands for the number of items, xi stands for the individual item and $\bar{x}$ for the mean across items for a given observation. After creating the variable, an independent sample t-test was conducted. Further analyses assessing the effect of the mediated moderation were performed using the PROCESS macro by Hayes (2012).

Four models using moderated mediation were used to test H4. The dependent variable was the created halo score for each model. The models differ in the mediator used. Model 1 through 4 respectively employ following mediators: Pos Affect, Neg Affect, Curiosity and Decentering. Induction was the independent variable and the description about the person (moral and neutral) serves as the moderator. Results are visible in Table 7.

Table 7

Results – Moderated mediation analysis of mindfulness and affect variables on halo score

	Model 1			**Model 2**			**Model 3**			**Model 4**		
Variables	**β**	***SE***	***p***	**β**	***SE***	***p***	**β**	***SE***	***p***	**β**	***SE***	***p***
Intercept	1.04	.17	**.00**	1.04	.18	**.00**	1.06	.18	.00	1.05	.17	**.00**
Induction Type	-.03	.25	.90	-.04	.25	.89	-.07	.26	.79	-.08	.24	.75
Pos Affect	-.01	.01	.48	-	-	-	-	-	-	-	-	-
Neg Affect	-	-	-	.02	.02	.35	-	-	-	-	-	-
Curiosity	-	-	-	-	-	-	.00	.02	.88	-	-	-
Decentering	-	-	-	-	-	-	-	-	-	.02	.02	.31
Person Description	.25	.25	.30	.34	.25	.18	.24	.26	.35	.17	.24	.47
Induction x Description	-.24	.37	.52	-.38	.37	.32	-.20	.39	.60	-.10	.36	.79
Mediator x Description	-.03	.02	.23	-.02	.04	.51	-.03	.03	.30	-.12	.04	**.00**
R^2	.06			.03			.04			.10		

*Outcome variable for all the models is halo score. Mediators are Curiosity, Decentering, Pos and Neg Affect. The different models employ different mediators in following order: Model 1 – Mediator: Positive Affect; Model 2 – Mediator: Negative Affect; Model 3 – Mediator: Curiosity; Model 4 – Mediator: Decentering Formatting indicates significance: $p < .05$ = 0, $p < .01$ = **0**, $p < .001$ = **0**

Three of the four moderated mediation models, namely Model 1 through 3, did not show significant results. The fourth model, which uses decentering as a mediator, revealed an interesting interaction with the description. Decentering alone did not significantly decrease or increase halo bias with a normal description, but when the person description was moral, higher levels of decentering significantly reduced halo bias. Furthermore, it is worth noting that person description did not significantly increase halo scores in any of the models. Meaning that a moral description did not lead to higher halo scores. Thus, H4 was partially supported.

5 Discussion

This chapter discusses the findings and practical as well as theoretical implications of the study. Furthermore, it explores the limitations of the study and highlights possible areas for future research.

5.1 Findings & Implications

Considering its methodology, this study contributed to the field of meditation by demonstrating that an induction, shorter than those applied in previous studies (see Hafenbrack, 2014; Hafenbrack 2022; Zhou et al., 2023), can still serve as an effective tool to increase mindfulness significantly. Furthermore, the positive effect of meditation on mood and affect was further supported through this study. Even a five-minute meditation improved mood, curiosity towards one's inner workings and decentering. This research also connects the literature of meditation and mindfulness to performance appraisals, two fields which have not been looked at jointly. Furthermore, this study provided partial support for the hypothesis that mindfulness can serve as a tool to counteract the halo bias. While the direct effect of the induction on halo bias was not statistically significant the moderated mediation model with decentering as a mediator showed significant results in the presence of a moral description. The mediating effect means that participants who were higher in decentering and additionally received the "moral" condition showed a clearly reduced halo bias compared to when a neutral description was provided. The reduction in this particular case was sensible as the positive description was supposed to enhance further halo effects, which in the "neutral" condition should have been less present. Curiosity, on the other hand, had no significant effect. However, the fact that the induction in the overall analysis did not play a significant role either suggests that parts of the mindfulness exercise have a different effect on the halo bias or that a singular session is not substantial enough to induce changes to detect a significant reduction. The findings suggest that state mindfulness, primarily through its channel decentering, may play a significant role in mitigating cognitive bias related to ratings.

From a practical and managerial standpoint, this thesis provides support for sensibility of short-term meditation for increasing mood of workers. Although some companies have adopted mindfulness into their structure (Gelles, 2015; Parcerisa, 2019), the programs that they offer are accompanied by high costs for the facilities, teachers and retreats they offer. Furthermore, loving-kindness, due to its focus on positive emotions and compassion seems to be suited to foster interpersonal relationships. This is also supported by findings such as increasing self-compassion (Reilly & Stuyvenberg, 2022), pro-social behaviour (Valor et al., 2024), and decreased negative interpersonal attitudes, which include

biases towards group, hatred, aggression and so forth (Zhou et al., 2023). Combined with the attributes of mindfulness to decrease judgmentalism and foster an open-hearted attitude (Feldman et al., 2007), which may foster a positive work-culture. Informal appraisals, such as feedback, which is given more frequently, could benefit from a more understanding and open-hearted approach. It may also contribute to improving people's communication skills (Chen et al., 2021), creating teams that can work together more effectively through clearer and more compassionate communication. Therefore, a meditation initiative in the workplace could lead to a more productive and fulfilling work-life as well as personal life, potentially leading to positively self-reinforcing loops.

5.2 Limitations

The areas of improvement and limitations of this study primarily concern the experimental setup and data analysis. In the following two sections, these topics will be explained thoroughly.

5.2.1 Experimental Setup

The areas of improvement and limitations of this study primarily concern the experimental setup. Novelty has been introduced mainly through the short duration of the induction (see Hafenbrack, 2013; Hafenbrack 2022; Zhou et al., 2023). The induction, though successful, could negatively influence the effect sizes and significance of findings due to its short duration. However, the short duration of the induction was a necessary imposition, since, as opposed to other studies in the field, most participants were collected through organic channels. The length of the study presented a significant challenge. In the initial phases of data collection, for example, the experiment was disseminated via the email distribution list of the Vienna University of Economics and Business, which contains over 30,000 individuals. This resulted in only 15 responses within two weeks and high abandonment rates. Thus, this indicates that participants' commitment to complete the study is low when not incentivised by monetary or any other non-monetary incentives.

The complexity of the experiment conducted in this study, including different media, led to difficulties for some participants. The level of complexity makes it more prone to misunderstandings and technical difficulties. Despite thorough efforts to create an ideal and user-friendly setup on Qualtrics, which included extensive testing prior to publication, some individuals reportedly experienced challenges. These most often included issues with either audio or video playback, rendering them unable to complete the study. No issues were found in Qualtrics, and any attempts to recreate the reported bugs and issues did not yield any results. Thus, it is still not explainable what caused difficulties for a select few people.

An additional point includes the environment in which the experiment is conducted. Initially, the study was conceptualised with the idea of being performed in a laboratory setting, as it would ideally require a controlled and quiet environment. Thus, by being unable to control for environmental factors, the study faces an additional challenge. Noises, interruptions, daily tasks, or multi-tasking, which, besides the implementation of additional tab-switching measures, cannot be tracked, can dilute the effects and lead to reduced effectiveness of the induction. Finally, the pace at which participants completed the experiment could not be controlled. Although extreme cases, where response duration exceeded a set amount of time, were removed, it is not possible to determine why some people took longer than expected. It could be due to several reasons. Someone may not pay attention, face difficulties, have the tab open in the background, and unknowingly start the timer, but effectively not commence with the experiment. Thus, the fidelity of engagement may not be given in a self-paced setting.

5.2.2 Data

The number of observations with 139 is deemed sufficient for the 2x2 experimental setup according to Brysbaert (2019) and exceeds the limit set by Simmons et al. (2011) of at least 20 observations per cell. Still, due to the moderation, more observations would enhance the validity of the study, especially considering the slight imbalance in group sizes after data cleaning. Additionally, due to the challenge of recruiting enough participants, 40 participants were recruited through Prolific. These participants were paid for participating. Each participant's response, collected through Prolific, was reviewed individually to ensure the quality of the answers. This included mainly checking click times for responses to assess whether somebody had listened to the induction or not, as well as looking for click patterns and monotonous responses that were indicative of poor quality. Four observations were rejected due to poor answer quality and removed from the final data set, and four did not pass the attention check. The bivariate analysis revealed no differences in variables related to the induction and affect. Thus, even though it is assumed that the removal of poorer quality responses did not compromise the study's integrity, it cannot be fully ruled out that the study suffers from variability in ratings due to participant motivation, attentiveness, and intention.

Lastly, participants were primarily students and young adults from Austria, which may limit the applicability of the study's findings to other contexts or countries. Through analysing the observations, possible ceiling effects could also be identified. Negative affect was significantly left-

skewed, with nearly no observations in the higher score ranges. Performance measures tended to be right-skewed, which means that for these variables, observations are limited to a narrower range, making it more difficult to find and measure significant differences depending on the treatment. Thus, especially for the performance measurement variable, a larger range that allows for more distinct assessments could lead to a more precise assessment of the effects of the treatments.

5.3 Future Research

Future research could expand on the mitigation of halo bias as well as other rater errors by focusing on longer and more controlled interventions. To be able to provide useful information on meditation from a more wholistic performance management point of view, the effects of short-form meditation could be assessed in more realistic settings, scenarios or case studies. Additionally, studying criteria such as people's satisfaction and performance, which is preferably tied to an objective measure, would further enhance the practical relevance of the studies. Further areas may focus on the interpersonal aspects of loving-kindness meditation and how increased compassion and potentially improved communication change team productivity as well as satisfaction.

6 Conclusion

Whether we like it or not, we are being rated and are rating things all around us all the time. We are ranking things in our minds, forming opinions about what we've eaten, watched, and with whom we've interacted. How we rank things and form our opinions is often influenced by our circumstances, the patterns of our thoughts, and we also tend to use mental shortcuts and biases (Kahneman & Tversky, 1974). Biases have been a persistent challenge of performance appraisals and have been documented scientifically as far back as 1890 (Edgeworth, 1890). A few of the most frequently recurring challenges in appraisals are rating biases (DeNisi & Murphy, 2017), such as the halo bias (Thorndike, 1920). Unsatisfactory results from approaches to improve appraisals even prompted researchers to call for a moratorium on performance ratings in general (DeNisi & Murphy, 2017; Landy & Farr, 1980). Regardless of moratoria being called, and workers and supervisors being displeased (Cleveland et al., 2007), ratings remained. Even the firms that allegedly abolished ratings mostly replaced them with a different form of rating, rather than removing ratings in general (Adler et al., 2016). Besides biases and inaccuracies, feedback can reduce self-esteem and perceived self-efficacy (Sargeant et al., 2008) and lead to negative emotional reactions (Lazarus, 1991). At the same time, meditation and mindfulness have been shown to increase positive emotions (Hafenbrack et al., 2020) and reduce biases (Maymin & Langer, 2021). Google, for example, has implemented several

courses and initiatives to promote smart work and a more positive work environment (Google, 2019). This thesis has provided support that even short periods of meditation can lead to significant increases in positive mood and awareness. Furthermore, partial support has been provided for the positive effect of mindfulness on the reduction of halo bias, which could lead to fairer and more accurate assessments. Mindfulness based interventions, like loving-kindness, could positively alter people's approach to appraisals, which could be a more sustainable approach than changing the tools with which they measure.

7 Sources

Aberdeen Group. 2010. Employee performance management: The alpha and the omega of talent strategy and business execution. Boston, MA: Aberdeen Group. Accessed at http://www.aberdeen.com/aberdeen-library/6361/RA-employee-performance-management.aspx.

Adler, S., Campion, M., Colquitt, A., Grubb, A., Murphy, K., Ollander-Krane, R., & Pulakos, E. D. (2016). Getting rid of performance ratings: Genius or folly? A debate. Industrial and Organizational Psychology, 9(2), 219-252.

Alexander, C. N., Robinson, P., Orme-Johnson, D. W., Schneider, R. H., et al. (1994). The effects of transcendental meditation compared to other methods of relaxation and meditation in reducing risk factors, morbidity, and mortality. *Homeostasis in Health and Disease, 35*(4-5), 243–263.

Alicke, M. D., Klotz, M. L., Breitenbecher, D. L., Yurak, T. J., & Vredenburg, D. S. (1995). Personal contact, individuation, and the better-than-average effect. *Journal of personality and social psychology*, *68*(5), 804.

Allen, D. G., & Griffeth, R. W. (2001). Test of a mediated performance–turnover relationship highlighting the moderating roles of visibility and reward contingency. *Journal of applied psychology*, *86*(5), 1014.

Armstrong, M. (2006). *Performance management 3rd ed*. Kogan Page Limited.

Armstrong, M and Baron, A , Managing performance: performance management in action New York Mc Graw Hill Book Co., 2005

Azeem, M. U., Bajwa, S. U., Shahzad, K., & Aslam, H. (2020). Psychological contract violation and turnover intention: The role of job dissatisfaction and work disengagement. Employee Relations: The International Journal, 42(6), 1291-1308. https://doi.org/10.1108/ER-09-2019-0372

Baer, R. A. (2003). Mindfulness training as a clinical intervention: A conceptual and empirical review. Clinical Psychology: Science and Practice, 10(2), 125–143. https://doi.org/10.1093/clipsy.bpg015

Baer, R. A., Smith, G. T., Hopkins, J., Krietemeyer, J., & Toney, L. (2006). Using self-report assessment methods to explore facets of mindfulness. *Assessment*, *13*(1), 27-45.

Baker, G., Gibbs, M., & Holmstrom, B. (1994). The wage policy of a firm. *The Quarterly Journal of Economics*, *109*(4), 921-955.

Balzer, W. K., & Sulsky, L. M. (1992). Halo and performance appraisal research: A critical examination. *Journal of Applied Psychology*, *77*(6), 975.

Bartlett, M. S. (1950). Tests of significance in factor analysis. British journal of psychology.

Benson, H., Beary, J. F., & Carol, M. P. (1974). The relaxation response. *Psychiatry*, *37*(1), 37-46.

Bernardin, H.J. and Russell, J.E. (1993) Human Resource Management: An Experiential Approach. McGraw-Hill, New York.

Bieri, J. (1966). Cognitive complexity and personality development. In *Experience structure & adaptability* (pp. 13-37). Berlin, Heidelberg: Springer Berlin Heidelberg.

Bishop, S. R., Lau, M., Shapiro, S., Carlson, L., Anderson, N. D., Carmody, J., Segal, Z. V., Abbey, S., Speca, M., Velting, D., & Devins, G. (2004). Mindfulness: A proposed operational definition. Clinical Psychology: Science and Practice, 11(3), 230–241. https://doi.org/10.1093/clipsy.bph077

Bland, J. M., & Altman, D. G. (1997). Statistics notes: Cronbach's alpha. *Bmj*, *314*(7080), 572.

Blankenship, K. L., Wegener, D. T., Petty, R. E., Detweiler-Bedell, B., & Macy, C. L. (2008). Elaboration and consequences of anchored estimates: An attitudinal perspective on numerical anchoring. Journal of Experimental Social Psychology, 44(6), 1465–1476. https://doi.org/10.1016/j.jesp.2008.07.005

Boswell, W. R., & Boudreau, J. W. (2002). Separating the developmental and evaluative performance appraisal uses. *Journal of Business and Psychology, 16*(3), 391–412. https://doi.org/10.1023/A:1012872907525

Bratton, Virginia K.; Dodd, Nancy G.; Brown, F. William (2011). The impact of emotional intelligence on accuracy of self-awareness and leadership performance. Leadership & Organization Development Journal, 32(2), 127–149. doi:10.1108/01437731111112971

Brockner, J., Grover, S. L., & Blonder, M. D. (1988). Predictors of survivors' job involvement following layoffs: A field study. *Journal of Applied Psychology, 73*(3), 436–442. https://doi.org/10.1037/0021-9010.73.3.436

Brown Michelle, Heywood John S. 2005. Performance appraisal systems: Determinants and change. 43(4): 659–79.

Brown, K. W., & Ryan, R. M. (2003). Mindful attention awareness scale. *Journal of personality and social psychology*.

Bunjak, A., Hafenbrack, A. C., Černe, M., & Arendt, J. F. W. (2022). Better to be optimistic, mindful, or both? The interaction between optimism, mindfulness, and task engagement. *Journal of Occupational and Organizational Psychology*, *95*(3), 595–623. https://doi.org/10.1111/joop.12389

Brumbach, G. B. (1988). Some issues, ideas and predictions about performance management. *Public Personnel Management, Pennsylvania Winter Press*.

Brysbaert, M. (2019). How many participants do we have to include in properly powered experiments? A tutorial of power analysis with reference tables. *Journal of cognition*, *2*(1), 16. doi: 10.5334/joc.72. PMID: 31517234; PMCID: PMC6640316.

Buckingham, M., & Goodall, A. (2015). Reinventing performance management. Harvard Business Review, 93(4), 40-50.

Burgoon, J. K., Berger, C. R., & Waldron, V. R. (2000). Mindfulness and interpersonal communication. Journal of Social Issues, 56(1), 105-127 https://doi.org/10.1111/0022-4537.00154

Cahn, B. R., & Polich, J. (2006). Meditation states and traits: EEG, ERP, and neuroimaging studies. *Psychological bulletin*, *132*(2), 180.

Carroll, S. J., & Schneier, C. E. (1982). Performance Appraisal and Review Systems: The Identification. *Measurement and Development of Performance in Organisations*.

Cason, G. J., & Cason, C. L. (1984). A deterministic theory of clinical performance rating: Promising early results. *Evaluation & the Health Professions*, *7*(2), 221-247.

Cappelli, P., & Conyon, M. J. (2017). What Do Performance Appraisals Do? ILR Review, 71(1), 88-116. https://doi.org/10.1177/0019793917698649 (Original work published 2018)

Cattell, J. M. (1906). American men of science: A biographical dictionary. New York, NY: Science Press.

Chen, H., Liu, C., Cao, X., Hong, B., Huang, D. H., Liu, C. Y., & Chiou, W. K. (2021). Effects of loving-kindness meditation on doctors' mindfulness, empathy, and communication skills. *International journal of environmental research and public health*, *18*(8), 4033.

Cleveland, J. N., Lim, A. S., & Murphy, K. R. (2007). Feedback phobia? Why employees do not want to give or receive performance feedback. In J. Langan-Fox, C. L. Cooper, & R. J. Klimoski (Eds.), Research companion to the dysfunctional workplace: Management challenges and symptoms (pp. 168–186). Edward Elgar Publishing. https://doi.org/10.4337/9781847207081.00018

Cleveland, J. N., Murphy, K. R., & Williams, R. E. (1989). Multiple uses of performance appraisal: Prevalence and correlates. Journal of applied psychology, 74(1), 130.

Coleman, J. W. (2002). *The new Buddhism: The western transformation of an ancient tradition*. Oxford University Press.

Conway, J. M. (1999). Distinguishing contextual performance from task performance for managerial jobs. Journal of applied Psychology, 84(1), 3.

Darley, J. M., & Fazio, R. H. (1980). Expectancy confirmation processes arising in the social interaction sequence. American Psychologist, 35(10), 867–881. https://doi.org/10.1037/0003-066X.35.10.867

Davids, T. W. R., & Stede, W. (Eds). (1959). The Pali Text Society's Pali-English dictionary. London: Luzac & Company, Ltd.

De Nisi, A. S., Randolph, W. A., & Blencoe, A. (1980, November). Peer evaluations: Causes and consequences. Proceedings of the 12th Annual Meeting of the American Institute of Decision Sciences (pp. 420-431). Las Vegas, NV.

Deming, W E (1986) Out of the Crisis, MIT Centre for Advanced Engineering Study, Cambridge, MA

Deming, W. E. (2018). *The new economics for industry, government, education*. MIT press.

DeNisi, A. S. (2000). Performance appraisal and performance management: a multilevel analysis. Multilevel theory, research and methods in organizations, 121-156.

DeNisi, A. S., & Murphy, K. R. (2017). Performance appraisal and performance management: 100 years of progress? Journal of Applied Psychology, 102(3), 421–433. https://doi.org/10.1037/apl0000085

Diener, E., Lucas, R. E., & Scollon, C. N. (2006). Beyond the hedonic treadmill: revising the adaptation theory of well-being. *American psychologist*, *61*(4), 305.

Dunnette, M. D., & Borman, W. C. (1979). Personnel selection and classification systems. Annual Review of Psychology, 30, 477–525. https://doi.org/10.1146/annurev.ps.30.020179.002401

Edgeworth, F.Y. 1890:The element of chance in competitive examinations. Journal of the Royal Statistical Society 53, 460-75 and 644-63.

Egan, G (1995) A clear path to peak performance, People Management, 18 May, pp 34–37

Erisman, S. M., & Roemer, L. (2010). A preliminary investigation of the effects of experimentally induced mindfulness on emotional responding to film clips. *Emotion*, *10*(1), 72–82. https://doi.org/10.1037/a0017162

Farias, M., Maraldi, E., Wallenkampf, K. C., & Lucchetti, G. (2020). Adverse events in meditation practices and meditation-based therapies: a systematic review. Acta Psychiatrica Scandinavica, 142(5), 374-393. https://doi.org/10.1111/acps.13225

Feldman, G., Hayes, A., Kumar, S., Greeson, J., & Laurenceau, J. P. (2007). Mindfulness and emotion regulation: The development and initial validation of the Cognitive and Affective Mindfulness Scale-Revised (CAMS-R). *Journal of psychopathology and Behavioral Assessment*, *29*(3), 177-190.

Fenton-O'Creevy Mark, Nicholson Nigel, Soane Emma, Willman Paul. 2005. Traders: Risks, Decisions and Management in Financial Markets. Oxford: Oxford University Press.

Ferris, G. R., Judge, T. A., Rowland, K. M., & Fitzgibbons, D. E. (1994). Subordinate influence and the performance evaluation process: Test of a model. Organizational behavior and human decision processes, 58(1), 101-135. https://doi.org/10.1006/obhd.1994.1030

FitMind. (2021, October 1). Guided Metta Meditation 5 Minutes | Sharon Salzberg [Video]. YouTube. https://www.youtube.com/watch?v=eCFfR5TgukI

Fisicaro, S. A., & Lance, C. E. (1990). Implications of three causal models for the measurement of halo error. Applied Psychological Measurement, 14, 419–429.

Flanagan, J. C. (1954). The critical incident technique. Psychological Bulletin, 51, 327–358. http://dx.doi.org/10.1037/h0061470

Fletcher, C (1993) Appraisal: Routes to improved performance, Institute of Personnel and Development, London

Fletcher, C. (2001) Performance Appraisal and Management. The Developing Research Agenda. Journal of Occupational and Organizational Psychology, 74, 473. https://doi.org/10.1348/096317901167488

Galante, J., Galante, I., Bekkers, M. J., & Gallacher, J. (2014). Effect of kindness-based meditation on health and well-being: a systematic review and meta-analysis. *Journal of consulting and clinical psychology*, *82*(6), 1101.

Galton, F., Hereditary Genius: An Inquiry Into Its Laws and Consequences. New York: Macmillan, 1892. Pp. xxvii, 368.

Gates, B. (2018, December 3). *The Headspace guide to meditation and mindfulness*. Gates Notes. Retrieved July 29, 2025, from https://www.gatesnotes.com/the-headspace-guide-to-meditation-and-mindfulness

Gelles, D. (2015). Mindful work: How meditation is changing business from the inside out. Houghton Mifflin Harcourt.

George, J. M., & Brief, A. P. (1992). Feeling good-doing good: A conceptual analysis of the mood at work-organizational spontaneity relationship. Psychological Bulletin, 112(2), 310–329. https://doi.org/10.1037/0033-2909.112.2.310

Germer, C. (2009). *The mindful path to self-compassion: Freeing yourself from destructive thoughts and emotions*. Guilford Press.

Gilbert, P., Basran, J., MacArthur, M., & Kirby, J. N. (2019). Differences in the semantics of prosocial words: an exploration of compassion and kindness. *Mindfulness*, *10*(11), 2259-2271.

Gino, F. (2008). Do we listen to advice just because we paid for it? The impact of advice cost on its use. *Organizational behavior and human decision processes*, *107*(2), 234-245.

Gist, M. E., & Mitchell, T. R. (1992). Self-efficacy: A theoretical analysis of its determinants and malleability. Academy of Management review, 17(2), 183-211. https://doi.org/10.5465/amr.1992.4279530

Goetz, J. L., & Simon-Thomas, E. (2017). The landscape of compassion: Definitions and scientific approaches. *The Oxford handbook of compassion science*, *1*, 3-15.

Gong, H., Ni, C. X., Liu, Y. Z., Zhang, Y., Su, W. J., Lian, Y. J., ... & Jiang, C. L. (2016). Mindfulness meditation for insomnia: A meta-analysis of randomized controlled trials. *Journal of psychosomatic research*, *89*, 1-6.

Google Trends. (n.d.). Medication and Zen [Online search interest data]. Retrieved July 27, 2025, from https://trends.google.com/trends/explore?date=today%205-y&q=meditation,zen&hl=en

Google. (2019, October 24). Mindfulness at work. Inside Google: Life at Google. Retrieved July 28, 2025, from https://blog.google/inside-google/life-at-google/mindfulness-at-work/

Gu, X., Luo, W., Zhao, X., Chen, Y., Zheng, Y., Zhou, J., ... & Li, R. (2022). The effects of loving-kindness and compassion meditation on life satisfaction: A systematic review and meta-analysis. *Applied Psychology: Health and Well-Being*, *14*(3), 1081-1101.

Guest, D E et al (1996) The State of the Psychological Contract in Employment, Institute of Personnel and Development, London

Hafenbrack, A. C., Cameron, L. D., Spreitzer, G. M., Zhang, C., Noval, L. J., & Shaffakat, S. (2020). Helping people by being in the present: Mindfulness increases prosocial behavior. Organizational Behavior and Human Decision Processes, 159, 21-38. https://doi.org/10.1016/j.obhdp.2019.08.005

Hafenbrack, A. C., Kinias, Z., & Barsade, S. G. (2014). Debiasing the mind through meditation: Mindfulness and the sunk-cost bias. *Psychological science*, *25*(2), 369-376.

Hafenbrack, A. C., LaPalme, M. L., & Solal, I. (2022). Mindfulness meditation reduces guilt and prosocial reparation. *Journal of Personality and Social Psychology*, *123*(1), 28.

Hall, B.J., Madigan, C., 2000. Compensation and performance evaluation at Arrow electronics. Harvard Business School Case, No. 9-800-290

Harari, O., & Zedeck, S. (1973). Development of behaviorally anchored scales for the evaluation of faculty teaching. *Journal of Applied Psychology*, *58*(2), 261–265. https://doi.org/10.1037/h0035633

Harrison, D. A., Virick, M., & William, S. (1996). Working without a net: Time, performance, and turnover under maximally contingent rewards. *Journal of Applied Psychology*, *81*(4), 331.

Hartshorne, H., May, M. A., & Maller, J. B. (1929). Studies in the nature of character, II Studies in service and self-control. MacMillan Co. https://doi.org/10.1037/11334-000

Harvey, N. (1997). Confidence in judgment. Trends in Cognitive Sciences, 1(2), 78–82. https://doi.org/10.1016/S1364-6613(97)01014-0

Harvey, P. (2012). *An introduction to Buddhism: Teachings, history and practices*. Cambridge University Press.

Hayes, A. F. (2012, February). *PROCESS: A versatile computational tool for observed variable mediation, moderation, and conditional process modeling*.

Headspace Inc. (2025). Headspace: Meditation & Sleep (Version current as of July 21, 2025) [Mobile application software]. Google Play Store. Retrieved July 29, 2025, from https://play.google.com/store/apps/details?id=com.getsomeheadspace.android&hl=en

Headspace. (n.d.). Headspace [Website homepage]. Retrieved July 27, 2025, from https://www.headspace.com

Hedge, J. W., & Kavanagh, M. J. (1988). Improving the accuracy of performance evaluations: Comparison of three methods of performance appraiser training. *Journal of Applied Psychology*, *73*, 68 –73. http://dx .doi.org/10.1037/0021-9010.73.1.68

Heneman, H. G., III. Comparisons of self- and superior ratings of managerial performance. Journal of Applied Psychology, 1974, 59, 638-642.

Holzbach, R. L. (1978). Rater bias in performance ratings: Superior, self-, and peer ratings. Journal of Applied Psychology, 63(5), 579–588. https://doi.org/10.1037/0021-9010.63.5.579

Houser, K. W. (2022). To Measure Is to Know … or Not. LEUKOS, 18(2), 103. https://doi.org/10.1080/15502724.2022.2029086

Hoyt, W. T. (2000). Rater bias in psychological research: When is it a problem and what can we do about it?. Psychological methods, 5(1), 64.

Hutcherson, C. A., Seppala, E. M., & Gross, J. J. (2008). Loving-kindness meditation increases social connectedness. *Emotion*, *8*(5), 720–724. https://doi.org/10.1037/a0013237

Ileana Petrescu, A. and Simmons, R. (2008), "Human resource management practices and workers' job satisfaction", International Journal of Manpower, Vol. 29 No. 7, pp. 651-667. https://doi.org/10.1108/01437720810908947

Jänig, W. (2003). The autonomic nervous system and its coordination by the brain. *Handbook of affective sciences,* 135-186.

Kabat-Zinn, J. (1982). An outpatient program in behavioral medicine for chronic pain patients based on the practice of mindfulness meditation: Theoretical considerations and preliminary results. General Hospital Psychiatry, 4, 33–47.

Kabat-Zinn, J. (1990). Full catastrophe living: Using the wisdom of your body and mind to face stress, pain and illness. New York: Delacorte.

Kabat-Zinn, J. (2003). Mindfulness-based interventions in context: Past, present, and future. Clinical Psychology: Science and Practice, 10(2), 144–156. https://doi.org/10.1093/clipsy.bpg016

Kabat-Zinn, J., Massion, M. D., Kristeller, J., Peterson, L. G., Fletcher, K. E., Pbert, L., et al. (1992). Effectiveness of a meditation-based stress reduction program in the treatment of anxiety disorders. American Journal of Psychiatry, 149, 936– 943.

Kaiser, H. F. (1974). An index of factorial simplicity. Psychometrika, 39(1), 31–36. https://doi.org/10.1007/BF02291575

Kampkötter, P. (2017). Performance appraisals and job satisfaction. *The International Journal of Human Resource Management*, *28*(5), 750-774.

Kane, J. S., Bernardin, H. J., Villanova, P., & Peyrefitte, J. (1995). Stability of rater leniency: Three studies. Academy of Management Journal, 38(4), 1036-1051. https://doi.org/10.5465/256619

Kassin, S. M., Goldstein, C. C., & Savitsky, K. (2003). Behavioral confirmation in the interrogation room: On the dangers of presuming guilt. Law and Human Behavior, 27(2), 187–203. https://doi.org/10.1023/A:1022599230598

Kearney, D. J., Malte, C. A., McManus, C., Martinez, M. E., Felleman, B., & Simpson, T. L. (2013). Loving-kindness meditation for posttraumatic stress disorder: A pilot study. *Journal of traumatic stress*, *26*(4), 426-434.

Kegan, R. (1982). *The evolving self: Problem and process in human development*. Harvard University Press.

Kirby, J. N., & Laczko, D. (2017). A randomized micro-trial of a loving-kindness meditation for young adults living at home with their parents. *Journal of Child and Family Studies*, *26*(7), 1888-1899.

Klimoski, R. J., & London, M. Role of the rater in performance appraisal. Journal of Applied Psychology, 1974, 59, 445-451.

Knauft, E. B. (1948). Construction and use of weighted check list rating scales for two industrial situations. Journal of Applied Psychology, 32, 63–70. http://dx.doi.org/10.1037/h0060388

Kneeland, N. (1929). That lenient tendency in rating. Personnel Journal, 7, 356-366.

Kucinskas, J. (2014). Change without confrontation: The making of mainstream meditation. Indiana University: Doctoral dissertation, Retrieved from http://pqdtopen.proquest.com/doc/1562698150.html?FMT=ABS.

Landy, F. J., & Farr, J. L. (1980). Performance rating. Psychological Bulletin, 87(1), 72–107. https://doi.org/10.1037/0033-2909.87.1.72

Landy, F. J., & Farr, J. L. (1983). The measurement of work performance. New York: Academic Press

Lau, M. A., Bishop, S. R., Segal, Z. V., Buis, T., Anderson, N. D., Carlson, L., ... & Devins, G. (2006). The Toronto mindfulness scale: Development and validation. Journal of clinical psychology, 62(12), 1445-1467. https://doi.org/10.1002/jclp.20326

Lawler, E E and McDermott, M (2003) Current performance management practices, WorldatWork Journal, Second quarter, pp 49–60

Lawler, E. E., III. The multitrait-multirater approach to measuring job performance. Journal of Applied Psychology, 1967, 51, 369-381.

Lazarus, R.S. (1991). Emotion and adaptation. New York: Oxford University Press. Lee, K., & Allen, N.J. (2002). Organizational citizenship behavior and workplace deviance: The role of affect and cognition. Journal of Applied Psychology, 87, 131–142

Lazear, E. P. (2000). Performance pay and productivity. *American Economic Review, 90(5),* 1346-1361.

Lefkowitz, J. (2000). The role of interpersonal affective regard in supervisory performance ratings: A literature review and proposed causal model. *Journal of Occupational and Organizational Psychology, 73(1)*, 67–85. https://doi.org/10.1348/096317900166886

Levine, J., & Butler, J. (1952). Lecture vs. group decision in changing behavior. J*ournal of Applied Psychology, 36*, 29 –33. http://dx.doi.org/ 10.1037/h0053624

Levy, P. E., & Williams, J. R. (2004). The Social Context for Performance Appraisal: A Review and Framework for the Future. *Journal of Management, 30*, 881-905. http://dx.doi.org/10.1016/j.jm.2004.06.005

Lewin, T. (2014, March 5). A New SAT Aims to Realign With Schoolwork. *The New York Times*. https://www.nytimes.com/2014/03/06/education/major-changes-in-sat-announced-by-college-board.html

Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1977). Calibration of probabilities: The state of the art. In *Decision Making and Change in Human Affairs: Proceedings of the Fifth Research Conference on Subjective Probability, Utility, and Decision Making, Darmstadt, 1–4 September, 1975* (pp. 275-324). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-010-1276-8_19

Lindahl, J.R. and Britton, W.B. (2019). 'I have this feeling of not really being here': Buddhist meditation and changes in sense of self. *Journal of Consciousness Studies, 26(7-8)*, 157-183.

Lockett, J (1992). Effective Performance Management, Kogan Page, London

Lumley, T., & McNamara, T. F. (1995). Rater characteristics and rater bias: implications for training. Language Testing, 12(1), 54-71. https://doi.org/10.1177/026553229501200104 (Original work published 1995)

Lv, J., Jiang, Y., Li, R., Chen, Y., Gu, X., Zhou, J., ... & Zeng, X. (2024). Effects of loving-kindness and compassion meditations on self-compassion: A systematic review and meta-analysis. *Clinical Psychology: Science and Practice*, *31*(1), 19.

Mani, B. G. (2002). Performance Appraisal Systems, Productivity, and Motivation: A Case Study. *Public Personnel Management, 31(2)*, 141-159. https://doi.org/10.1177/009102600203100202 (Original work published 2002)

May, C. J., Johnson, K., & Weyker, J. R. (2016). Experimental approaches to loving-kindness meditation and mindfulness that bridge the gap between clinicians and researchers. In *Mindfulness and Buddhist-derived approaches in mental health and addiction* (pp. 85-93). Cham: Springer International Publishing.

Maymin, P. Z., & Langer, E. J. (2021). Cognitive biases and mindfulness. *Humanities and Social Sciences Communications, 8(1)*, 1-11.

McGee, M. (2008). Meditation and psychiatry. *Psychiatry (Edgmont)*, *5*(1), 28.

Menefee, J. A., & Murphy, R. O. (2004). Rewarding and retaining the best: Compensation strategies for top performers. *Benefits Quarterly, 20(3)*, 13.

Meyer, H. H. (1975). The pay-for-performance dilemma. *Organizational Dynamics, 3*, 39-50.

Meyer, H. H., Kay, E., & French, J. R. P. (1965). Split roles in performance appraisal. *Harvard Business Review, 43 (1),* 123-129.

Miceli, M. P., Jung, I., Near, J. P., & Greenberger, D. B. (1991). Predictors and outcomes of reactions to pay-for-performance plans. *Journal of Applied Psychology, 76(4),* 508–521. https://doi.org/10.1037/0021-9010.76.4.508

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

Motowidlo, S. J., & Van Scotter, J. R. (1994). Evidence that task performance should be distinguished from contextual performance. *Journal of Applied Psychology, 79(4)*, 475–480. https://doi.org/10.1037/0021-9010.79.4.475

Murphy, K. R., & Cleveland, J. N. (1995). *Understanding performance appraisal: Social, organizational, and goal-based perspectives*. Sage.

Murphy, K. R., Cleveland, J. N., Skattebo, A. L., & Kinney, T. B. (2004). Raters who pursue different goals give different ratings. *Journal of Applied Psychology, 89(1*), 158–164. https://doi.org/10.1037/0021-9010.89.1.158

Neff, K. (2003). Self-compassion: An alternative conceptualization of a healthy attitude toward oneself. *Self and identity*, *2*(2), 85-101.

Newcomb, T. M. (1929). *The consistency of certain extrovert-introvert behavior patterns in 51 problem boys* (No. 382). Teachers College, Columbia University.

Newcomb, T. An experiment designed to test the validity of a rating technique. Journal of Educational Psychology, 1931, 22, 279-289.

Odean, T. (1998). Are investors reluctant to realize their losses?. *The Journal of finance*, *53*(5), 1775-1798.

Ortlieb, S. A., Kügel, W. A., & Carbon, C. C. (2020). Fechner (1866): The aesthetic association principle—A commented translation. *i-Perception*, *11*(3), 2041669520920309.

Ospina, M. B. (2008). Meditation practices for health: state of the research.

Palmer, J. K., & Loveland, J. M. (2008). The influence of group discussion on performance judgments: Rating accuracy, contrast effects, and halo. The Journal of Psychology, 142(2), 117-130.

Parcerisa, C. (2019, October 24). Can mindfulness actually help you work smarter? Google Blog. Retrieved July 29, 2025, from https://blog.google/inside-google/life-at-google/mindfulness-at-work/

Patel, P. M., & Deshpande, V. A. (2017). Application of plan-do-check-act cycle for quality and productivity improvement-A review. *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*, *5*(1), 197-201.

Paterson, D. G. (1922). The Scott Company Graphic Rating Scale. Journal of Personnel Research, 1, 361–376.

Pearce, J. L., & Porter, L. W. (1986). Employee responses to formal performance appraisal feedback. *Journal of applied psychology*, *71*(2), 211.

Perry, J. L., & Porter, L. W. (1981). Organizational assessments of the Civil Service Reform Act of 1978. Washington, D.C.: U.S. Office of Personnel Management.

Phillips, J. M., & Gully, S. M. (1997). Role of goal orientation, ability, need for achievement, and locus of control in the self-efficacy and goal--setting process. Journal of Applied Psychology, 82(5), 792–802. https://doi.org/10.1037/0021-9010.82.5.792

Pohl, R. F., & Hell, W. (1996). No reduction in hindsight bias after complete information and repeated testing. *Organizational Behavior and Human Decision Processes*, *67*(1), 49-58.

Prendergast, C. (1999). The provision of incentives in firms. *Journal of economic literature, 37(1)*, 7-63.

Pulakos, E. D., & O'Leary, R. S. (2011). Why is performance management broken? Industrial and Organizational Psychology: Perspectives on Science and Practice, 4(2), 146–164. https://doi.org/10.1111/j.1754-9434.2011.01315.x

Randall, J., & Sim, A. (2013). *Managing people at work*. Routledge.

Raoelison, Matthieu, and Wim De Neys. "Do we de-bias ourselves?: The impact of repeated presentation on the bat-and-ball problem." *Judgment and Decision making* 14, no. 2 (2019): 170-178.

Rapgay, L., & Bystrisky, A. (2009). Classical mindfulness: An introduction to its theory and practice for clinical application. *Annals of the New York Academy of Sciences*, *1172*(1), 148-162.

Raubenheimer, A. S., An Experimental Study of Some Behavior Traits of the Potentially Delinquent Boy. Psychol. Rev., Monog. Suppl, 1925, 34, pp. 107.

Realyvásquez-Vargas, A., Arredondo-Soto, K. C., Carrillo-Gutiérrez, T., & Ravelo, G. (2018). Applying the Plan-Do-Check-Act (PDCA) cycle to reduce the defects in the manufacturing industry. A case study. *Applied Sciences*, *8*(11), 2181.

Ream, M. J. (1921). A statistical method for incomplete ordering of merit ratings. Journal of Applied Psychology, 5, 261–266. http://dx.doi.org/10 .1037/h0066886

Reilly, E. B., & Stuyvenberg, C. L. (2023). A meta-analysis of loving-kindness meditations on self-compassion. *Mindfulness*, *14*(10), 2299-2310.

Roch, S. G., Woehr, D. J., Mishra, V., & Kieszczynska, U. (2012). Rater training revisited: An updated meta-analytic review of frame-of-reference training. *Journal of occupational and organizational psychology*, *85*(2), 370-395.

Ross, J., & Staw, B. M. (1993). Organizational escalation and exit: Lessons from the Shoreham nuclear power plant. *Academy of Management journal*, *36*(4), 701-732.

Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In Advances in experimental social psychology (Vol. 10, pp. 173-220). Academic Press. https://doi.org/10.1016/S0065-2601(08)60357-3

Rugg, H. (1921). Is the rating of human character practicable? Journal of Educational Psychology, 12, 425– 438, 485–501.

Salzberg, S., & Kabat-zinn, J. (2004). *Lovingkindness: The Revolutionary Art of Happiness*. Shambhala Publications.

Sargeant, J., Mann, K., Sinclair, D., Van der Vleuten, C., & Metsemakers, J. (2008). Understanding the influence of emotions and reflection upon multi-source feedback acceptance and use. Advances in Health Sciences Education, 13(3), 275-288.

Sedlmeier, P., Eberth, J., Schwarz, M., Zimmermann, D., Haarig, F., Jaeger, S., & Kunze, S. (2012). The psychological effects of meditation: a meta-analysis. *Psychological bulletin, 138*(6), 1139.

Schneier, C. E. (1977). Operational utility and psychometric characteristics of Behavioral Expectation Scales: A cognitive reinterpretation. Journal of Applied Psychology, 62(5), 541–548. https://doi.org/10.1037/0021-9010.62.5.541

Shapiro, D. H. (1992). A preliminary study of long term meditators: Goals, effects, religious orientation, cognitions. *Journal of Transpersonal Psychology*, *24*(1), 23-39.

Shapiro, S. L., Carlson, L. E., Astin, J. A., & Freedman, B. (2006). Mechanisms of mindfulness. *Journal of clinical psychology*, *62*(3), 373-386.

Shapiro, S. L., & Schwartz, G. E. (2000). Intentional systemic mindfulness: an integrative model for self-regulation and health. Advances in mind-body medicine, 16(2), 128-134. doi: 10.1054/ambm.1999.0118. PMID: 10835770.

Sharon, A. T., & Bartlett, C. J. (1969). Effect of instructional conditions in producing leniency on two types of rating scales. Personnel Psychology, 22(3), 251–263. https://doi.org/10.1111/j.1744-6570.1969.tb00330.x

Shear, J. (2006). *The experience of meditation: Experts introduce the major traditions*. Paragon House.

Siegel, R.D., Germer, C.K., Olendzki, A. (2009). Mindfulness: What Is It? Where Did It Come From?. In: Didonna, F. (eds) Clinical Handbook of Mindfulness. Springer, New York, NY. https://doi.org/10.1007/978-0-387-09593-6_2

Sim, M., Kim, S.-Y., & Suh, Y. (2021). Sample Size Requirements for Simple and Complex Mediation Models. Educational and Psychological Measurement, 82(1), 76-106. https://doi.org/10.1177/00131644211003261

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological science*, *22*(11), 1359-1366. https://doi.org/10.1177/0956797611417632 (Original work published 2011)

Singh, N. N., Wahler, R. G., Adkins, A. D., Myers, R. E., & Mindfulness Research Group. (2003). Soles of the feet: A mindfulness-based self-control intervention for aggression by an individual with mild mental retardation and mental illness. *Research in Developmental Disabilities*, *24*(3), 158-169. doi: 10.1016/s0891-4222(03)00026-x. PMID: 12742385.

Sivanathan, N., & Cynthia Fekken, G. (2002). Emotional intelligence, moral reasoning and transformational leadership. *Leadership & Organization Development Journal*, *23*(4), 198-204.

Smith, P. C., & Kendall, L. M. (1963). Retranlsation of expectations: An approach to the construction of unambiguous anchors for rating scales. Journal of Applied Psychology, 47, 149 –155. http://dx.doi.org/10.1037/ h0047060

Spence, J. R., & Keeping, L. (2011). Conscious rating distortion in performance appraisal: A review, commentary, and proposed framework for research. Human Resource Management Review, 21(2), 85–95. https://doi.org/10.1016/j.hrmr.2010.09.013

Statista. (n.d.). Gen Z who were often anxious or depressed by age, U.S. [Statistics]. Retrieved July 29, 2025, from https://www.statista.com/statistics/1613301/gen-z-often-anxious-depressed-by-age-us/

Stellar, J. E., & Willer, R. (2018). Unethical and inept? The influence of moral information on perceptions of competence. *Journal of Personality and Social Psychology*, *114*(2), 195.

Sugi, Y., & Akutsu, K. (1968). Studies on respiration and energy-metabolism during sittiing in zazen. *Taiikugaku kenkyu (Japan Journal of Physical Education, Health and Sport Sciences)*, *12*(3), 190-206.

Symonds P. M., Notes on Rating: Rating versus Ranking. /. Appl. Psychol, 1925, 9, 188-195.

Taylor, M. S., Fisher, C. D., & ugen, D. R. (1984). Individuals' reactions to performance feedback in organizations: A control theory perspective. In K. M. Rowland & G. R. Ferris (Eds.), Research in personnel and human resources management (pp. 81-124). Greenwich, CN: JAI Press

Tedeschi, J. T., & Melburg, V. (1984). Impression management and influence in the organization. Research in the Sociology of Organizations, 3, 31–58.

Terman Lewis M., Trait Ratings. [In] Genetic Studies of Genius. Vol. I, Mental and Physical Traits of a Thousand Gifted Children. Stanford: Stanford University Press, 1925. Pp. xv, 648 (pp. 518-555).

Thompson, E. R. (2007). Development and validation of an internationally reliable short-form of the positive and negative affect schedule (PANAS). *Journal of cross-cultural psychology*, *38*(2), 227-242.

Thorndike, E. L. (1920). A constant error in psychological ratings. Journal of Applied Psychology, 4, 25–29. http://dx.doi.org/10.1037/h0071663

Thornton, G. C. (1968). The relationship between supervisory-and self-appraisals of executive performance. Personnel Psychology, 21(4), 441-455. https://doi.org/10.1111/j.1744-6570.1968.tb02044.x

Trank, C. Q., Rynes, S. L., & Bretz Jr, R. D. (2002). Attracting applicants in the war for talent: Differences in work preferences among high achievers. Journal of business and psychology, 16(3), 331-345.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. Cognitive psychology, 5(2), 207-232. https://doi.org/10.1016/0010-0285(73)90033-9

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. Science, 185(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124

VandeWalle, D., & Cummings, L. L. (1997). A test of the influence of goal orientation on the feedback-seeking process. Journal of Applied Psychology, 82(3), 390–400. https://doi.org/10.1037/0021-9010.82.3.390

Valor, C., Martínez-de-Ibarreta, C., Carrero, I., & Merino, A. (2024). Effects of loving-kindness meditation on prosocial behavior: empirical and meta-analytic evidence. *Journal of Social Marketing, 14(2),* 280-300.

Walsh, R., & Shapiro, S. L. (2006). The meeting of meditative disciplines and Western psychology: a mutually enriching dialogue. *American psychologist*, *61*(3), 227.

Walters, M. E. (1995). The performance management handbook.

Watson, D., Clark, L. A., & Carey, G. (1988). Positive and negative affectivity and their relation to anxiety and depressive disorders. *Journal of abnormal psychology*, *97*(3), 346.

Wayne, S. J., & Ferris, G. R. (1990). Influence tactics, affect, and exchange quality in supervisor-subordinate interactions: A laboratory experiment and field study. Journal of Applied Psychology, 75(5), 487–499. https://doi.org/10.1037/0021-9010.75.5.487

Wayne, S. J., & Liden, R. C. (1995). Effects of impression management on performance ratings: A longitudinal study. *Academy of Management journal*, *38*(1), 232-260.

Weiss, L. A. (1933). Rating scales with special references to the field of child development. Psychological Bulletin, 30(3), 185–208. https://doi.org/10.1037/h0075716

Wells, F. L. (1907). A Statistical Study of Literary Merit. (Columbia Univ. Cont. to Phil. & Psych., 16, 3.). *Archives of Psychology*.

West, M. (1979). Meditation. *The British Journal of Psychiatry*, *135*(5), 457-467.

West, M. A. (2016). The practice of meditation. *The psychology of meditation: Research and practice*, 3-25.

Winstanley, D and Stuart-Smith, K (1996) Policing performance: the ethics of performance management, Personnel Review, 25 (6), pp 66–84

Woehr, D. J. (1994). Understanding frame-of-reference training: The impact of training on the recall of performance information. *Journal of applied psychology*, *79*(4), 525.

Woehr, D. J., & Huffcutt, A. I. (1994). Rater training for performance appraisal: A quantitative review. *Journal of occupational and organizational psychology*, *67*(3), 189-205.

Woolfolk, R. L. (1975). Psychophysiological correlates of meditation. *Archives of General Psychiatry*, *32*(10), 1326-1333.

Wright, G. (1982). Changes in the realism and distribution of probability assessments as a function of question type. Acta Psychologica, 52(1), 165–174. https://doi.org/10.1016/0001-6918(82)90033-6

Young, M., & Dulewicz, V. (2007). Relationships between emotional and congruent self-awareness and performance in the British Royal Navy. *Journal of Managerial Psychology*, *22*(5), 465-478.

Zeng, X., Chiu, C. P., Wang, R., Oei, T. P., & Leung, F. Y. (2015). The effect of loving-kindness meditation on positive emotions: A meta-analytic review. *Frontiers in psychology*, *6*, 1693.

Zhao, C., Liu, Y., & Gao, Z. (2016). An identification perspective of servant leadership's effects. *Journal of Managerial Psychology*, *31*(5), 898-913.

Zhou, J., Lang, Y., Wang, Z., Gao, C., Lv, J., Zheng, Y., ... & Zeng, X. (2023). A meta-analysis and systematic review of the effect of loving-kindness and compassion meditations on negative interpersonal attitudes. *Current Psychology, 42(31)*, 27813-27827.

8 Appendix

8.1 Links & Materials

Video link: Game Theory 101: Iterated Elimination of Strictly Dominated Strategies
Excerpt taken from this recoding: NY-Times Recording
Guided meditation link (only audio was played): Guided Metta Meditation 5 Minutes | Sharon Salzberg

8.2 Experiment

Start of Block: Description

Dear participant, Thank you for your time and willingness to contribute to this survey! My name is Daniel Alacs and I am carrying out this study as part of my Master's thesis research within the scope of decision making at WU Vienna University of Economics and Business and Católica Lisbon School of Business & Economics. The survey should **take around 12 to 15 minutes** to complete. You will listen to an audio recording and watch a short video. The audio recording is five minutes long and the video recording is two minutes long. There are no right or wrong answers, so please answer honestly based on your personal opinion. This study is completely anonymous, and all information provided will be kept strictly confidential. All collected data will be used solely for academic research purposes. If you have any questions please feel free to contact me: h11911419@s.wu.ac.at or at s-dalacs@ucp.pt. **Thank you! Important**: To correctly answer this survey, we kindly ask you to **wear headphones or make sure you can play sound out loud**. Please make sure to always **scroll down to see the "next" button**. You will be asked to listen to an audio recording and shown a short video during the study. Also, your contribution to the study is voluntary. You can decide to break off the survey at any time by closing the tab.

End of Block: Description

Start of Block: Demographics

What is your gender?

- ○ Male (1)
- ○ Female (2)
- ○ Other (3) _____
- ○ Prefer not to say (4)

How old are you?

_____

What is your country of origin?

▼ Afghanistan (1) ... Zimbabwe (196)

What best describes your occupation over the last three months?

- ○ Employed full time (1)
- ○ Employed part time (2)
- ○ Unemployed looking for work (3)
- ○ Unemployed not looking for work (4)
- ○ Retired (5)
- ○ Student (6)
- ○ Other (7) _____

JS

Imagine that this ladder scale pictures how society is set up. **At the top of the ladder are the people who are best off** – they have the most money, the highest amount of schooling, and the jobs that bring the most respect. **At the bottom are people who are the worst off** – they have the least money, little or no education, no job or jobs that no one wants or respects. Now think about your family. Please tell us where you think your family would be on this scale. (you must click on the gray area of the scale and not drag/slide the scale if you are on a touchscreen)

The people who are the best off

The people who are the worst off

You think your family would be a : _____ (1)

Total : _____

End of Block: Demographics

Start of Block: Audio test

Please **make sure to always scroll to the bottom** of the page to see the n**ext button.**

Please listen to the audio above and select the word that has been mentioned. This is an attention check.

- ○ Orange (1)
- ○ Lemon (2)
- ○ Apple (3)
- ○ Strawberry (4)

End of Block: Audio test

Start of Block: Induction - LMK

JS

Please listen carefully to this audio recording. This is a guided meditation about spreading kindness and love. Make sure you have headphones on or are in a room where you can play the audio in an undisturbed fashion. Try to relax and follow the instructions of the voice and participate in the guided meditation. You will be able to click "next" after the audio has finished.

Timing
First Click (1)
Last Click (2)
Page Submit (3)
Click Count (4)

End of Block: Induction - LMK

Start of Block: NY-Times

JS

Please listen carefully to this audio about the SAT exam. We will ask some questions about your experience later. You will be able to click "next" after the audio has finished.

Timing
First Click (1)
Last Click (2)
Page Submit (3)
Click Count (4)

End of Block: NY-Times

Start of Block: Manipulation Check - simple question

Please indicate which statement best describes the audio you just listened to. This is an attention check.

- ○ The audio focused on practicing slow and steady breathing. (1)
- ○ The audio discussed proposed changes to college classes. (2)
- ○ The audio emphasized spreading kindness and love. (3)
- ○ The audio talked about making adjustments to the SAT exam. (4)

End of Block: Manipulation Check - simple question

Start of Block: TMS

We would now like to ask you some questions regarding your experience of listening to the previous audio. Please state how accurately these statements describe your experience while listening to the previous audio recording.

I experienced myself as separate from my changing thoughts and feelings.

	Not at all	A little	Moderately	Quite a bit	Very much
	0	1	2	3	4
1 ()					

I was rather open to my experiences and feelings than trying to controll or change them.

	Not at all	A little	Moderately	Quite a bit	Very much
	0	1	2	3	4
1 ()					

I was curious about what I might learn about myself by taking notice of how I react to certain thoughts, feelings, or sensations.

	Not at all	A little	Moderately	Quite a bit	Very much
	0	1	2	3	4
1 ()					

I experienced my thoughts more as events in my mind than as a necessarily accurate reflection of the way things 'really' are.

	Not at all	A little	Moderately	Quite a bit	Very much
	0	1	2	3	4
1 ()					

I was curious to see what my mind was up to from moment to moment.

	Not at all	A little	Moderately	Quite a bit	Very much
	0	1	2	3	4
1 ()					

I was curious about each of the thoughts and feelings that I was having.

	Not at all	A little	Moderately	Quite a bit	Very much
	0	1	2	3	4
1 ()					

I was receptive to observing unpleasant thoughts and feelings without interfering with them.

	Not at all	A little	Moderately	Quite a bit	Very much
	0	1	2	3	4
1 ()					

I was more invested in just watching my experiences as they arose, than in figuring out what they could mean.

	Not at all	A little	Moderately	Quite a bit	Very much
	0	1	2	3	4

1 ()	

I approached each experience by trying to accept it, no matter whether it was pleasant or unpleasant.

Not at all	A little	Moderately	Quite a bit	Very much
0	1	2	3	4

1 ()	

I remained curious about the nature of each experience as it arose.

Not at all	A little	Moderately	Quite a bit	Very much
0	1	2	3	4

1 ()	

I was aware of my thoughts and feelings without overidentifying with them.

Not at all	A little	Moderately	Quite a bit	Very much

	0	1	2	3	4
1 ()					

I was curious about my reactions to things.

	Not at all	A little	Moderately	Quite a bit	Very much
	0	1	2	3	4
1 ()					

I was curious about what I might learn about myself by just taking notice of what my attention gets drawn to.

	Not at all	A little	Moderately	Quite a bit	Very much
	0	1	2	3	4
1 ()					

End of Block: TMS

Start of Block: PANAS-SF

For each of these emotions indicate how you feel at the present moment

	Very slightly or not at all (1)	A little (2)	Moderately (3)	Quite a bit (4)	Extremely (5)
Interested (1)	○	○	○	○	○
Distressed (2)	○	○	○	○	○
Excited (3)	○	○	○	○	○
Upset (4)	○	○	○	○	○
Strong (5)	○	○	○	○	○
Guilty (6)	○	○	○	○	○
Scared (7)	○	○	○	○	○
Hostile (8)	○	○	○	○	○
Enthusiastic (9)	○	○	○	○	○
Proud (10)	○	○	○	○	○
Irritable (11)	○	○	○	○	○
Alert (12)	○	○	○	○	○

Ashamed (13)	○	○	○	○	○
Inspired (14)	○	○	○	○	○
Nervous (15)	○	○	○	○	○
Determined (16)	○	○	○	○	○
Attentive (17)	○	○	○	○	○
Jittery (18)	○	○	○	○	○
Active (19)	○	○	○	○	○
Afraid (20)	○	○	○	○	○

End of Block: PANAS-SF

Start of Block: Description of Person - Neutral

Imagine you are a student and have to take a class on game theory. Your professor is a 28-year-old male who is currently writing his Ph.D. at the University of Wisconsin. He was born in Columbus, Ohio, USA. Furthermore, the student magazine of his university reports that he usually goes for walks around the university campus after lunch. He also owns a car and takes good care of it. The car is a family heirloom and he makes sure that the car is in great condition despite the fact that it has suffered some damage over the years. Also, he does not get involved in wars between his co-workers, which are particularly fierce in his workplace. He knows that choosing a side could be an advantage, but prefers to stay out of it. You can find a picture of this man below. You will be presented with 2 minute a video recording of this man and are supposed to rate him on several factors regarding your opinion of his teaching capabilities.

End of Block: Description of Person - Neutral

Start of Block: Description of Person - Moral

Imagine you are a student and have to take a class on game theory. Your professor is a 28-year-old male who is currently writing his Ph.D. at the University of Wisconsin and was born in Columbus, Ohio, USA. Furthermore, the student magazine of his university reports that he has donated a large amount of money to a charity for homeless kids. He also adopted a dog and is a very caring owner. The dog has special needs (it is blind) and he makes sure that the dog lives a happy and fulfilling life despite the fact that the dog can't see. Also, he stopped the spread of a false rumour about a coworker, which said she was cheating on her husband. He knew the rumour would hurt his coworker and that it was not true. You can see a picture of this man below. You will be presented with a 2 minute video recording of this man and are supposed to rate him on several factors regarding your opinion of his teaching capabilities.

End of Block: Description of Person - Moral

Start of Block: Video

JS

Watch this two and a half minute video of our fictitious person teaching one of the introductory topics to his course. Pay attention carefully, as you must rate him based on his performance. (You might need to scroll down to see the continue button. It will only appear at the end of the video.) (Make sure when you watch the video so that the audio track is in English. You can switch audiotrack by clicking on the settings in the lower right corner of the video.)

Timing
First Click (1)
Last Click (2)
Page Submit (3)
Click Count (4)

End of Block: Video

Start of Block: Rating

Please rate the professor in the dimensions below. Rate him according to your opinion of his teaching capabilities based on the video you have just watched.

Delivery - This refers to the use of speech, tone of voice, pacing to effectively communicate a topic.

	Very poor					Excellent
	1	2	3	4	5	6	7
()						

Depth of knowledge - This refers to how knowledgeable a person is in their field and how thoroughly they are able to explain topics.

	Very Poor					Excellent
	1	2	3	4	5	6	7
()						

Structuredness of the lecture - This refers to the clarity of the agenda, smoothnes of transitions between topics and general coherence.

	Very poor					Excellent

	1	2	3	4	5	6	7
()							

Relations with students - This refers to how approachable, welcoming and respectful a person is when engaging with students.

	Very poor						Excellent
	1	2	3	4	5	6	7
()							

Fairness & grading - This refers to transparency of grades and thoroughness and timeliness of feedback

	Very poor						Excellent
	1	2	3	4	5	6	7
()							

End of Block: Rating

8.3 Factor Analysis – Loading Tables

Table 8

TMS items factor loadings

	Loadings
Variables	**MR1**
TMS 1	.38
TMS 2	.51
TMS 3	.74
TMS 4	.50
TMS 5	.73
TMS 6	.81
TMS 7	.52
TMS 8	.45
TMS 9	.51
TMS 10	.79
TMS 11	.30
TMS 12	.78
TMS 13	.78

Table 9

Performance items factor loadings

	Loadings	
Variables	**MR1**	**MR2**
1: Delivery	.73	.25
2: Knowledge	.68	.10
3: Structure	.55	.31
4: Relations	.56	.48
5: Grading	.22	1.00

Table 10

PANAS items factor loadings

	Loadings	
Variables	**MR1**	**MR2**
1: Interested	.65	-
2: Distressed	-	.51
3: Excited	.80	-
4: Upset	-	.69
5: Strong	.64	-
6: Guilty	-	.50
7: Scared	-	.71
8: Hostile	-	.43
9: Enthusiastic	.82	-
10: Proud	.74	-
11: Irritable	-	.56
12: Alert	.43	.23
13: Ashamed	-	.73
14: Inspired	.78	-
15: Nervous	-	.75
16: Determined	.73	-
17: Attentive	.63	-
18: Jittery	.16	.73
19: Active	.69	-
20: Afraid	-	.77

8.4 PROCESS R-Ouptut

8.4.1 Model 1

********** PROCESS Procedure for R Version 5.0 **********

Written by Andrew F. Hayes, Ph.D. www.afhayes.com
Documentation available in Hayes (2022). www.guilford.com/p/hayes3

****
Model: 15
Y: halo_score
X: induction_type
M: positive_affect
W: person_descr
Sample size: 139
Custom seed: 654321
****
Outcome Variable: positive_affect
Model Summary:

R	R-sq	MSE	F	df1	df2	p
0.2188	0.0479	70.2804	6.8877	1.0000	137.0000	0.0097

Model:

	coeff	se	t	p	LLCI	ULCI
constant	-1.6991	0.9616	-1.7669	0.0795	-3.6006	0.2025
induction_type	3.7487	1.4284	2.6245	0.0097	0.9242	6.5733

****

Outcome Variable: halo_score

Model Summary:

R	R-sq	MSE	F	df1	df2	p
0.2559	0.0655	1.1136	1.8642	5.0000	133.0000	0.1048

Model:

	coeff	se	t	p	LLCI	ULCI
constant	1.0462	0.1742	6.0068	0.0000	0.7017	1.3907
induction_type	-0.0293	0.2532	-0.1159	0.9079	-0.5302	0.4716
positive_affect	-0.0104	0.0145	-0.7157	0.4754	-0.0391	0.0183
person_descr	0.2525	0.2455	1.0287	0.3055	-0.2330	0.7380
int_1	-0.2375	0.3701	-0.6418	0.5221	-0.9696	0.4945
int_2	-0.0260	0.0216	-1.2016	0.2317	-0.0688	0.0168

Product terms key:

int_1 : induction_type x person_descr

int_2 : positive_affect x person_descr

Test(s) of highest order unconditional interaction(s):

	R2-chng	F	df1	df2	p
X*W	0.0029	0.4119	1.0000	133.0000	0.5221
M*W	0.0101	1.4438	1.0000	133.0000	0.2317

********** DIRECT AND INDIRECT EFFECTS OF X ON Y **********

Conditional direct effect(s) of X on Y:

person_descr	effect	se	t	p	LLCI	ULCI
0.0000	-0.0293	0.2532	-0.1159	0.9079	-0.5302	0.4716
1.0000	-0.2669	0.2699	-0.9887	0.3246	-0.8008	0.2670

Conditional indirect effects of X on Y:

INDIRECT EFFECT:

induction_type -> positive_affect -> halo_score

person_descr	Effect	BootSE	BootLLCI	BootULCI
0.0000	-0.0390	0.0557	-0.1472	0.0811
1. 0000	-0.1365	0.1089	-0.4026	0.0173

Index of moderated mediation

(differences beween conditional indirect effects):

	Index	BootSE	BootLLCI	BootULCI
person_descr	-0.0975	0.1199	-0.4006	0.0775

********** BOOTSTRAP RESULTS FOR REGRESSION MODEL PARAMETERS **********

Outcome variable: positive_affect

	Coeff	BootMean	BootSE	BootLLCI	BootULCI
constant	-1.6991	-1.7002	0.9474	-3.5218	0.1673
induction_type	3.7487	3.7567	1.4269	0.9081	6.5494

Outcome variable: halo_score

	Coeff	BootMean	BootSE	BootLLCI	BootULCI
constant	1.0462	1.0487	0.1412	0.7885	1.3484
induction_type	-0.0293	-0.0343	0.2216	-0.4701	0.4040
positive_affect	-0.0104	-0.0101	0.0142	-0.0376	0.0180
person_descr	0.2525	0.2427	0.2421	-0.2295	0.7201
int_1	-0.2375	-0.2282	0.3320	-0.8738	0.4301
int_2	-0.0260	-0.0263	0.0264	-0.0795	0.0243

********** ANALYSIS NOTES AND ERRORS **********

Level of confidence for all confidence intervals in output: 95

Number of bootstraps for percentile bootstrap confidence intervals: 10000

NOTE: The following variables were mean centered prior to analysis:

positive_affect

8.4.2 Model 2

Model: 15

Y: halo_score

X: induction_type

M: negative_affect

W: person_descr

Sample size: 139

Custom seed: 654321

****

Outcome Variable: negative_affect

Model Summary:

R	R-sq	MSE	F	df1	df2	p
0.0433	0.0019	28.2510	0.2579	1.0000	137.0000	0.6124

Model:

	coeff	se	t	p	LLCI	ULCI
constant	-0.2084	0.6097	-0.3419	0.7330	-1.4141	0.9972
induction_type	0.4599	0.9056	0.5078	0.6124	-1.3309	2.2507

****

Outcome Variable: halo_score

Model Summary:

R	R-sq	MSE	F	df1	df2	p
0.1797	0.0323	1.1532	0.8879	5.0000	133.0000	0.4912

Model:

	coeff	se	t	p	LLCI	ULCI
constant	1.0432	0.1772	5.8878	0.0000	0.6927	1.3936
induction_type	-0.0367	0.2547	-0.1440	0.8857	-0.5405	0.4672
negative_affect	0.0208	0.0219	0.9460	0.3459	-0.0226	0.0642
person_descr	0.3360	0.2488	1.3505	0.1791	-0.1561	0.8281
int_1	-0.3754	0.3746	-1.0021	0.3181	-1.1163	0.3655
int_2	-0.0240	0.0364	-0.6589	0.5111	-0.0961	0.0481

Product terms key:

int_1 : induction_type x person_descr

int_2 : negative_affect x person_descr

Test(s) of highest order unconditional interaction(s):

	R2-chng	F	df1	df2	p
X*W	0.0073	1.0043	1.0000	133.0000	0.3181
M*W	0.0032	0.4341	1.0000	133.0000	0.5111

********** DIRECT AND INDIRECT EFFECTS OF X ON Y **********

Conditional direct effect(s) of X on Y:

person_descr	effect	se	t	p	LLCI	ULCI
0.0000	-0.0367	0.2547	-0.1440	0.8857	-0.5405	0.4672
1. 0000	-0.4121	0.2746	-1.5004	0.1359	-0.9553	0.1312

Conditional indirect effects of X on Y:

INDIRECT EFFECT:

induction_type -> negative_affect -> halo_score

person_descr	Effect	BootSE	BootLLCI	BootULCI
0.0000	0.0095	0.0299	-0.0342	0.0861
1. 0000	-0.0015	0.0249	-0.0535	0.0543

Index of moderated mediation

(differences beween conditional indirect effects):

	Index	BootSE	BootLLCI	BootULCI
person_descr	-0.0110	0.0399	-0.1085	0.0611

********** BOOTSTRAP RESULTS FOR REGRESSION MODEL PARAMETERS **********

Outcome variable: negative_affect

	Coeff	BootMean	BootSE	BootLLCI	BootULCI
constant	-0.2084	-0.1982	0.5802	-1.2559	1.0080
induction_type	0.4599	0.4593	0.8979	-1.2888	2.2810

Outcome variable: halo_score

	Coeff	BootMean	BootSE	BootLLCI	BootULCI
constant	1.0432	1.0455	0.1417	0.7834	1.3382
induction_type	-0.0367	-0.0329	0.2133	-0.4433	0.3833
negative_affect	0.0208	0.0230	0.0226	-0.0129	0.0752
person_descr	0.3360	0.3393	0.2687	-0.1735	0.8821
int_1	-0.3754	-0.3869	0.3406	-1.0577	0.2822
int_2	-0.0240	-0.0249	0.0356	-0.0996	0.0438

********** ANALYSIS NOTES AND ERRORS **********

Level of confidence for all confidence intervals in output: 95

Number of bootstraps for percentile bootstrap confidence intervals: 10000

NOTE: The following variables were mean centered prior to analysis:

negative_affect

8.4.3 Model 3

Model: 15

Y: halo_score

X: induction_type

M: curiosity

W: person_descr

Sample size: 139

Custom seed: 654321

****

Outcome Variable: curiosity

Model Summary:

R	R-sq	MSE	F	df1	df2	p
0.2991	0.0895	33.7519	13.4612	1.0000	137.0000	0.0003

Model:

	coeff	se	t	p	LLCI	ULCI
constant	-1.6461	0.6664	-2.4700	0.0147	-2.9638	-0.3283
induction_type	3.6318	0.9899	3.6689	0.0003	1.6744	5.5892

****

Outcome Variable: halo_score

Model Summary:

R	R-sq	MSE	F	df1	df2	p
0.1924	0.0370	1.1475	1.0228	5.0000	133.0000	0.4069

Model:

	coeff	se	t	p	LLCI	ULCI
constant	1.0587	0.1764	6.0024	0.0000	0.7098	1.4076
induction_type	-0.0705	0.2579	-0.2732	0.7851	-0.5806	0.4396
curiosity	0.0030	0.0207	0.1431	0.8864	-0.0379	0.0438
person_descr	0.2382	0.2555	0.9322	0.3529	-0.2672	0.7436
int_1	-0.2042	0.3880	-0.5264	0.5995	-0.9717	0.5632
int_2	-0.0340	0.0324	-1.0488	0.2962	-0.0982	0.0301

Product terms key:

int_1 : induction_type x person_descr

int_2 : curiosity x person_descr

Test(s) of highest order unconditional interaction(s):

	R2-chng	F	df1	df2	p
X*W	0.0020	0.2771	1.0000	133.0000	0.5995
M*W	0.0080	1.1000	1.0000	133.0000	0.2962

********** DIRECT AND INDIRECT EFFECTS OF X ON Y **********

Conditional direct effect(s) of X on Y:

person_descr	effect	se	t	p	LLCI	ULCI
0.0000	-0.0705	0.2579	-0.2732	0.7851	-0.5806	0.4396
1.0000	-0.2747	0.2899	-0.9477	0.3450	-0.8481	0.2987

Conditional indirect effects of X on Y:

INDIRECT EFFECT:

induction_type -> curiosity -> halo_score

person_descr	Effect	BootSE	BootLLCI	BootULCI
0.0000	0.0107	0.0751	-0.1295	0.1753
1.0000	-0.1128	0.1316	-0.4077	0.1158

Index of moderated mediation

(differences beween conditional indirect effects):

	Index	BootSE	BootLLCI	BootULCI
person_descr	-0.1235	0.1538	-0.4742	0.1369

********** BOOTSTRAP RESULTS FOR REGRESSION MODEL PARAMETERS **********

Outcome variable: curiosity

	Coeff	BootMean	BootSE	BootLLCI	BootULCI
constant	-1.6461	-1.6456	0.6854	-2.9649	-0.2837
induction_type	3.6318	3.6148	0.9761	1.6828	5.5285

Outcome variable: halo_score

	Coeff	BootMean	BootSE	BootLLCI	BootULCI
constant	1.0587	1.0611	0.1458	0.7874	1.3615
induction_type	-0.0705	-0.0730	0.2368	-0.5401	0.3908
curiosity	0.0030	0.0033	0.0195	-0.0357	0.0413
person_descr	0.2382	0.2318	0.2449	-0.2433	0.7200
int_1	-0.2042	-0.2019	0.3370	-0.8652	0.4527
int_2	-0.0340	-0.0345	0.0390	-0.1128	0.0396

********** ANALYSIS NOTES AND ERRORS **********

Level of confidence for all confidence intervals in output: 95

Number of bootstraps for percentile bootstrap confidence intervals: 10000

NOTE: The following variables were mean centered prior to analysis:

curiosity

8.4.4 Model 4

Model: 15

Y: halo_score

X: induction_type

M: decentering

W: person_descr

Sample size: 139

Custom seed: 13181

****

Outcome Variable: decentering

Model Summary:

R	R-sq	MSE	F	df1	df2	p
0.1665	0.0277	24.2313	3.9063	1.0000	137.0000	0.0501

Model:

	coeff	se	t	p	LLCI	ULCI
constant	-0.7513	0.5647	-1.3306	0.1855	-1.8679	0.3652
induction_type	1.6577	0.8387	1.9764	0.0501	-0.0008	3.3162

****

Outcome Variable: halo_score

Model Summary:

R	R-sq	MSE	F	df1	df2	p
0.3303	0.1091	1.0617	3.2566	5.0000	133.0000	0.0083

Model:

	coeff	se	t	p	LLCI	ULCI
constant	1.0510	0.1695	6.2004	0.0000	0.7157	1.3863
induction_type	-0.0781	0.2434	-0.3210	0.7487	-0.5596	0.4033
decentering	0.0245	0.0241	1.0158	0.3116	-0.0232	0.0722
person_descr	0.1738	0.2411	0.7208	0.4723	-0.3031	0.6507
int_1	-0.0964	0.3601	-0.2677	0.7893	-0.8086	0.6158
int_2	-0.1169	0.0365	-3.2059	0.0017	-0.1890	-0.0448

Product terms key:

int_1 : induction_type x person_descr

int_2 : decentering x person_descr

Test(s) of highest order unconditional interaction(s):

	R2-chng	F	df1	df2	p
X*W	0.0005	0.0717	1.0000	133.0000	0.7893
M*W	0.0688	10.2776	1.0000	133.0000	0.0017

Focal predictor: decentering (M)

Moderator: person_descr (W)

Conditional effects of the focal predictor at values of the moderator(s):

person_descr	effect	se	t	p	LLCI	ULCI
0.0000	0.0245	0.0241	1.0158	0.3116	-0.0232	0.0722
1.0000	-0.0924	0.0273	-3.3786	0.0010	-0.1465	-0.0383

********** DIRECT AND INDIRECT EFFECTS OF X ON Y **********

Conditional direct effect(s) of X on Y:

person_descr	effect	se	t	p	LLCI	ULCI
0.0000	-0.0781	0.2434	-0.3210	0.7487	-0.5596	0.4033
1.0000	-0.1745	0.2653	-0.6578	0.5118	-0.6994	0.3503

Conditional indirect effects of X on Y:

INDIRECT EFFECT:

induction_type -> decentering -> halo_score

person_descr	Effect	BootSE	BootLLCI	BootULCI
0.0000	0.0406	0.0379	-0.0271	0.1226
1.0000	-0.1531	0.1223	-0.4391	0.0219

Index of moderated mediation

(differences beween conditional indirect effects):

	Index	BootSE	BootLLCI	BootULCI
person_descr	-0.1937	0.1377	-0.5107	0.0121

********** BOOTSTRAP RESULTS FOR REGRESSION MODEL PARAMETERS **********

Outcome variable: decentering

	Coeff	BootMean	BootSE	BootLLCI	BootULCI
constant	-0.7513	-0.7527	0.6162	-1.9630	0.4460
induction_type	1.6577	1.6585	0.8192	0.0580	3.2656

Outcome variable: halo_score

	Coeff	BootMean	BootSE	BootLLCI	BootULCI
constant	1.0510	1.0497	0.1397	0.7932	1.3383
induction_type	-0.0781	-0.0723	0.2115	-0.4835	0.3487
decentering	0.0245	0.0241	0.0189	-0.0136	0.0610
person_descr	0.1738	0.1670	0.2145	-0.2551	0.5842
int_1	-0.0964	-0.0971	0.2976	-0.6944	0.4817
int_2	-0.1169	-0.1133	0.0521	-0.2106	-0.0088

********** ANALYSIS NOTES AND ERRORS **********

Level of confidence for all confidence intervals in output: 95

Number of bootstraps for percentile bootstrap confidence intervals: 10000

NOTE: The following variables were mean centered prior to analysis:

decentering