Published Biannually by
The Department of Counseling and Clinical Psychology
Teachers College, Columbia University

Fall 2025 – Volume 25

Graduate Student Journal of Psychology

About the Artist:

George Bonanno is a Professor of Clinical Psychology at Columbia University's Teachers College and internationally recognized for his pioneering research on human resilience in the face of loss and potential trauma. He is recognized by the Web of Science as among the top one percent most cited scientists in the world, and has been honored with lifetime achievement awards by the Association for Psychological Science (APS), the International Society for Traumatic Stress Studies (ISTSS), and the International Positive Psychology Association (IPPA). In addition to the books, The End of Trauma and The Other Side of Sadness, Dr. Bonanno has published hundreds of peer-reviewed scientific articles, many appearing in leading journals such as Nature, JAMA, American Psychologist, and the Annual Review of Psychology. He is also an avid painter (when he has time), reads widely, and loves music.

I began drawing around the age of 15. If I remember correctly, one day I simply decided to try a small portrait drawing. The process was thrilling. Hours passed and I hardly noticed it. Ever since that day I have passionately engaged in portraiture, landscapes and still lifes in a variety of media. I've never sought formal training in art but nonetheless hoped to make a career of it. Oddly my biggest artistic output occurred while I was a doctoral student in psychology, often painting and drawing in the middle of the night. I exhibited and sold a number of my works during that time. Yet, simultaneously, I found psychological research and writing deeply gratifying, and eventually that became the obvious career choice. I never regretted that choice but, luckily, when time permits, I am still able to lose myself in the wonders of creating art.

- Dr. George Bonnano

FALL 2025 – VOLUME 25

Graduate Student Journal of Psychology

The Graduate Student Journal of Psychology is a publication of the Department of Counseling and Clinical Psychology, Teachers College, Columbia University.

AIMS AND SCOPE. Founded in 1998, the Graduate Student Journal of Psychology (GSJP) is a peer-reviewed journal consisting of empirical and theoretical articles drawing from all areas of academic psychology. The GSJP is published biannually and includes original research articles, case studies, theoretical articles, and brief reports in the areas of adult and child psychopathology, the psychotherapeutic process, community mental health, multiculturalism, assessment, child development, ethics, professional development, and other topics with a relationship to research or practice in psychology.

MANUSCRIPTS. Manuscript submission for Volume 25 will be handled on a rolling basis. Submit manuscripts according to the submission guidelines printed in this issue. Please email us your manuscript along with a cover letter to gsjp@tc.columbia.edu. For more information on submission process visit www.tc.edu/publications/gsjp.

FALL 2025 – VOLUME 25

Graduate Student Journal of Psychology

FOREWORD

ARTICLES

Letter from the Editors:

We are pleased to introduce Volume 25 of the Graduate Student Journal of Psychology. This edition showcases rigorous scholarship that advances psychological science and addresses critical challenges in our field.

This volume features six compelling articles spanning diverse methodologies and substantive areas. Clayton offers a novel conceptual model for healthcare practitioners navigating weight-related conversations with patients. Drawing on motivational interviewing principles, the model promotes behavior change while reducing weight stigma, a timely contribution given the documented harms of weight bias in healthcare settings. Khan et al. examine relationships among attachment styles, mentalization, and suicidal ideation in a Pakistani sample, underscoring the importance of emotion regulation strategies in therapy with anxiously attached individuals. Zhu and White employ the Minimal Group Paradigm to investigate competing influences of ingroup membership and physical attractiveness on decision-making, revealing the robust primacy of group identity.

Additional contributions include Esterine's longitudinal investigation of prenatal anxiety and child language development at 18 months, which raises important questions about developmental mechanisms and timing; OConnell et al.'s methodological comparison of data quality across MTurk and Prolific, providing evidence-based guidance for researchers; and Zentner and Yildirim-Erbasli's innovative application of response time analysis and machine learning to detect careless responding in distressed survey participants, achieving strong predictive accuracy with support vector machines.

This edition marks an exciting transition for our team. We welcome three new associate editors—Madeleine, Kendall, and Joey—whose dedication and fresh perspectives have already enriched this volume and will carry forward our tradition of scholarly excellence.

We extend deep gratitude to our contributors, peer reviewers, and faculty sponsor, Dr. Matt Blanchard, for their invaluable support. We also thank our outgoing editors, whose leadership continues to inspire our work.

We invite you to engage with the research presented in this volume and join the conversation. Connect with us at gsjp@tc.columbia.edu or through our social media platforms. Thank you for your continued support as we advance the field together.

Warm regards,

The Editors

Graduate Student Journal of Psychology
Fall 2025 - Vol. 25

A Conceptual Model for Addressing Weight Stigma and Health: Guiding Practitioner Conversations for Weight Health

Colter K. Clayton, MA
Department of Psychology, University of Mississippi, USA

Practitioners working in a variety of healthcare settings increasingly face a dilemma when speaking with patients about weight health. On one hand, prescriptive weight-related health advice can exacerbate stigma, while on the other, ignoring insufficient health behavior engagement limits health and increases the risk of other adverse weight-related health conditions. Research has demonstrated that higher-than-optimal body weight is a correlate of morbidity and mortality, but has also demonstrated that weight stigma is pervasive, negatively impacting health, health behavior, and well-being. This article introduces a novel conceptual model to help practitioners initiate conversations about weight health by striving to support health behavior change in a way that deactivates and disempowers weight stigma. By advancing the acceptance principle from motivational interviewing and adapting its scope, the model focuses on destigmatizing attitudes and assumptions related to weight health to prevent or reduce generalized and internalized weight stigma. The model also focuses on limiting interpersonal stigma and its disruptive role in practitioner-patient communication by supporting personal autonomy for a lifestyle of health behavior. This article reports results from a rapid review and calls for research efforts to examine the potential causal role of active acceptance for reducing weight stigma. Overall, the conceptual model simultaneously promotes health behavior and reduces weight stigma for weight health.
Keywords: weight stigma, weight health, health behavior, motivational interviewing, healthcare practitioner

People don't care how much you know until they know how much you care.

— Attributed to Theodore Roosevelt

People need support. Every person faces adversity, and health is a universal concern – whether young or old, rich or poor, all contend with morbidity and mortality. Compassionate individuals are empowered by a deep desire to promote the well-being of others. Whether they work in medicine, health psychology, or public health, their passion for care sparks career interests and fuels professional diligence. However, how can healthcare practitioners (some researchers prefer *practitioner to provider*; Scarff, 2021) from distinct training backgrounds and professional roles navigate the complexities of weight-related health and weight stigma? This paper reconciles promoting health behavior and reducing stigma for weight health by proposing a novel conceptual model for practitioners. The model resolves a dilemma faced by many practitioners: saying too much (e.g., prescriptive weight-related health advice exacerbating stigma) versus too little (e.g., ignoring insufficient health behavior engagement). To fully articulate the model, it is first necessary to review the evidence demonstrating the importance of weight-related health and discuss weight stigma as a barrier to health.

Weight Health

Weight-related health is a principal public health concern globally (Okunogbe et al., 2022; Safaei et al., 2021). Being of higher body weight due to an excess of adiposity large enough to cause reduced health or longevity – at least probabilistically over time – is linked to morbidity and mortality, is largely considered a complex lifestyle concern, and its consequences are associated with increased risk for a myriad of cardiometabolic diseases (Allison et al., 2008; Fruh, 2017; Safaei et al., 2021). Indeed, higher-than-optimal body weight has been linked to increased risk, or at least concomitant to, "... nearly every chronic condition, from diabetes, to dyslipidemia, to poor mental health. Its impacts on risk of stroke and cardiovascular disease, certain cancers, and osteoarthritis are significant" (Hruby & Hu, 2015, p. 10). Whether certain ranges of body weight (e.g., body mass index > 30; Allison et al., 2008; Katz, 2014) are to be medically considered a disease is irrelevant to the development of the model. The strength of evidence demonstrates that it is beyond a reasonable doubt that being of higher body weight due to excessive accumulation of adipose tissue is related to poorer cardiometabolic health and well-being (Hruby & Hu, 2015; Pi-Sunyer, 2009; Robinson et al., 2020; Visscher & Seidell, 2001). Although weight-related health and associated conditions are complex in their etiology and maintenance, consequences related to morbidity and mortality are considered largely preventable or reducible (Fruh, 2017; Hruby & Hu, 2015). Prominent health organizations educate practitioners and patients about the negative impacts of higher-than-optimal body weight. The World Health Organization, Centers for Disease Control and Prevention, and the

National Institutes of Health use websites[1*] to publicly disseminate the latest research on risk factors and best-practice health strategies to educate practitioners and patients.

Healthcare practitioners know that weight loss (when needed) has a large impact on improving health through lowering risk for disease (e.g., type 2 diabetes, hypertension, dyslipidemia; Haase et al., 2021) and is linked to improvement in current health conditions (Pojednic et al., 2022). Even modest amounts of weight loss (e.g., >5% of body weight) are associated with clinically significant improvements in cardiometabolic health and emotional well-being (Fruh, 2017), and weight loss at higher levels can lead to further improvements (Ryan & Yockey, 2017). More engagement in health behaviors is needed; physical activity, eating a healthy diet, and obtaining quality sleep are vital for health. For example, in the United States, only about 1 in 4 adults meets physical activity guidelines, and engagement is even lower among those with lower incomes (Elgaddal et al., 2020).

Practitioners also know that physical activity has health benefits independent of body weight. For example, among people with higher body weight, physical activity has demonstrated positive changes at the cellular level (e.g., improved enzyme function), along with improved metabolic function (e.g., increased insulin sensitivity), and better cardiovascular outcomes (e.g., decreased arterial stiffness; Pojednic et al., 2022). Thus, physiologic improvements observed with increased physical activity are observed despite starting weight and irrespective of whether weight loss occurs, and cardiorespiratory fitness is an indicator of metabolic health (Ortega et al., 2013). Given that health behaviors such as physical activity contribute to health, increased adherence and implementation are needed.

Weight Stigma

Practitioners are increasingly becoming aware of how health stigma and discrimination limit social acceptance and opportunity, and how they exacerbate inequality, which are well-documented barriers to engagement in healthcare and health behaviors (Hatzenbuehler et al., 2013; Stangl et al., 2019). Weight stigma is a form of health stigma. It refers to reduced social status due to excess body weight, and weight-related negative attitudes are linked to discriminatory actions such as unfair treatment in healthcare (Rubino et al., 2020). A landmark study at the turn of the century by pioneering weight stigma researchers introduced its negative impact on employment, education, and healthcare (Puhl & Brownell, 2001). Subsequent research has solidified weight stigma as a major health issue: at least 1 in 2 adults experience stigma related to their weight, which is associated with less regular medical checkups, healthcare avoidance, and worse healthcare quality experiences (Puhl et al., 2021). Recent international research initiatives to address weight stigma have called for more actions from the medical community to promote the education and training of healthcare professionals to practice without participation in or perpetuation of weight stigma (Puhl et al., 2021; Puhl, 2023; Rubino et al., 2020).

Healthcare practitioners are increasingly recognizing that weight stigma is not helpful or healthy. A major concern of weight stigma is its negative effect on health behaviors: decreased physical activity, decreased healthcare engagement, and paradoxically, even weight gain (Tomiyama, 2014). Weight stigma has negative impacts in laboratory and real-life contexts (Major et al., 2014; Panza et al., 2023; Puhl & Suh, 2015; Rubino et al., 2020). For example, brief exposure to reading or watching weight stigma can elicit cardiovascular reactivity and increase overeating behavior (Major et al., 2014; Panza et al., 2023). The everyday consequences of weight stigma include maladaptive eating behaviors (e.g., binge eating), lower engagement in physical activity, and unhealthy weight gain (Puhl & Suh, 2015).

Dilemmas of Weight Health and Stigma

Saying too much can exacerbate stigma. For example, some researchers have called for increased social pressures (e.g., societal shame) to help individuals become more aware of their weight and associated stigmatization (Callahan, 2013), but subsequent stigma research did not support social pressure as a viable method to improve weight health (Puhl et al., 2021). Conversely, saying too little may condone poor health behaviors. The Health at Every Size (HAES) approach is a product of both academic research and social movements that challenge weight stigma, deprioritize weight and weight-related biomarkers of health, and

1 *https://www.who.int/health-topics/obesity#tab=tab_1; https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight; https://www.cdc.gov/family-healthy-weight/php/recognized-programs/index.html; https://www.cdc.gov/obesity/php/about/obesity-strategies-what-can-be-done.html; https://www.niddk.nih.gov/health-information/weight-management/adult-overweight-obesity/health-risks

promote bodily acceptance to improve health and well-being (Bombak, 2014; Penney & Kirk, 2015). Facilitating access to healthcare that is free of shame is a prominent virtue of this framework. However, while intending to reduce stigma, dismissing (whether deliberately or inadvertently) physical health recommendations may occur.

People with higher weight report negative psychological reactions to receiving weight-related advice from healthcare practitioners (Standen et al., 2024), which may explain why some groups reject and even criticize weight health recommendations as unethical (e.g., HAES). Self-affirmation theory describes how people maintain integrity to themselves (Steele, 1988). In weight stigma research, this theory may explain why people may be motivated to maintain a positive self-view regarding their weight when perceiving weight stigma. For example, a healthcare professional recommending weight loss may evoke a patient's efforts to maintain self-integrity, where healthcare recommendations related to body weight are perceived as attacks on personal identity; this may be a reason that recommendations can fail to produce motivation for change. Thus, practitioners must simultaneously preserve patients' self-integrity and recommend health behavior change.

Given the interacting complexities of weight-related health and stigma, practitioners face a daunting responsibility to care for patients. The dilemma practitioners face is two competing risks: over-alerting patients can be counterproductive, eliciting societal or personal shame and exacerbating the distress of stigma; conversely, under-alerting patients to weight-related health risks fails to promote positive health behavior change. A paradigm grounded in theory and evidence is needed to guide practitioners as they navigate the complexities of weight-related health.

Developing the WHISTLE Model

Research on the biopsychosocial factors of weight-related health and healthcare has led to theoretical and conceptual models that are comprehensive and often complex (Marks, 2015; McCabe et al., 2023; Michie et al., 2014; Plotnikoff et al., 2007). However, such models are not specific to stigma, though one model of weight stigma has emphasized a vicious feedback loop of stigma and weight gain (Tomiyama, 2014). Thus, existing models fall short of offering a simple and practical framework to begin helpful discussions about weight health and stigma. The novel Weight Health through Integrated Stigma-Reduction and Lifestyle Engagement (WHISTLE) model serves as a guide for practitioners. In contrast to the traditional "sick" role of passive prescriptions, contemporary healthcare increasingly prioritizes enhancing patient agency to boost health outcomes through more effective health behavior (Armstrong, 2014). However, healthcare practitioners may feel uncomfortable or incompetent in discussing weight with patients (Pont et al., 2017). The WHISTLE model aids practitioners as they practice attitudes of acceptance to reduce weight stigma when opening a conversation about weight-related health that facilitates the encouragement of health behavior change by using a patient-centered approach.

Health behavior change is facilitated through the atheoretical principles of motivational interviewing. At its core, motivational interviewing is an empirically supported, patient-centered approach to talking with people in a way that strengthens their own motivation and commitment for behavioral change (Miller & Rollnick, 2023). Reflections and questions by practitioners help patients work through the natural ambivalence they experience when facing decisions about behavioral change. The proposed mechanism that has received the most empirical support is practitioners' use of selectively reinforcing patients' own motivational statements for behavior change (Bischof et al., 2021). Research on motivational interviewing as a favorable intervention for positive health behavior (e.g., physical activity) and weight-related health outcomes is mixed (Amiri et al., 2022; Frost et al., 2018; Lundahl et al., 2013; Makin et al., 2021; Michalopoulou et al., 2022). However, motivational interviewing as an empirically supported way of effectively being in a helping relationship (e.g., practitioner engagement) and facilitating behavior change has received substantial empirical support across healthcare settings (Bischof et al., 2021; Lundahl et al., 2013; Luty & Iwanowicz, 2018; Magill et al., 2018; Miller & Rollnick, 2023; Rubak et al., 2005).

While motivational interviewing principles broadly support behavioral change, they do not address the issue of stigma – particularly weight-related stigma – in weight-related health outcomes. The novel WHISTLE model is a person-centered framework that targets weight health conversations and was derived from a component known as acceptance. The WHISTLE model's guiding principle of acceptance was based on a previous edition of the motivational interviewing approach created by Miller & Rollnick (2013) with four 'A's: absolute worth, accurate empathy, affirmation, and autonomy support. In the WHISTLE model, the four 'A's function as fundamental assumptions, attitudes, and actions for practitioners working with patients and weight health concerns. Thus, the WHISTILE model provides a foundation to simultaneously reduce stigma and encourage health behavior change.

In the WHISTLE model, practitioners assume and adopt an attitude that each patient has inherent worth irrespective of weight, health, or behavior, and that each person is innately capable of behavioral change. Actions of practitioner acceptance include expressing accurate empathy and communicating a genuine interest in understanding the patient's experiences and situation (Schumacher & Madson, 2014). Practitioners also provide affirmations and statements to help patients see their own strengths and resources and champion patients' autonomy as they choose whether and how to make behavioral change. Since patients with a history of weight stigma report being less heard and respected by their healthcare practitioner (Puhl et al., 2021), the WHISTLE model's assumptions, attitudes, and actions function to reduce self-stigma, stigma in the practitioner-patient relationship, and prevent or reduce generalized weight stigma from interfering with healthcare discussions about weight health (see Figure 1).

Weight stigma interferes with practitioner-patient communication through shameful language and impaired emotional engagement, which are associated with healthcare disengagement and avoidance (Puhl, 2023). Practitioners with higher weight stigmatizing attitudes report feeling less confident in offering weight health recommendations and less likely to use a person-centered approach when speaking with patients (Bennett & Puhl, 2024). On the other hand, within a motivational interviewing approach, patients who talk about plans for behavior change show better body weight outcomes over time (Copeland et al., 2017). Patients can feel simultaneously motivated to lose weight and feel bad (e.g., guilty) regarding their weight-related health (Standen et al., 2024), highlighting the need for concurrent stigma reduction and health behavior encouragement. Practitioners may be better equipped to work with weight health to deactivate stigma with acceptance-promoting statements.

The WHISTLE model helps practitioners to channel an atmosphere of acceptance to 1) prevent generalized social stigma and their own potentially stigmatizing attitude from entering the relationship, and 2) reduce self-stigmatizing beliefs among patients (though the patient acts as the mediator of this relationship; see Figure 1). Within this atmosphere, practicing acceptance moves generalized practitioner stigma and self-stigma further away from the patient to allow for increased patient self-efficacy, weight-related health information discussions, and encouragement of healthy lifestyle behavior. The figurative cloud represents the subjective, relational aspect in which practitioner-patient communication occurs. Statements of accurate empathy increase connection and empower patients' autonomy for making behavioral health changes (see Table 1). Supporting autonomy is intended to have the dual effect of further reducing stigma while promoting behavior change. Practitioners risk eliciting ambivalent or negative psychological reactions as they offer weight-related advice, but a two-way discussion is a characteristic of patients feeling motivated to begin healthy behavior change (Standen et al., 2024). Practitioners using the WHISTLE model who create an atmosphere for health behavior change, champion patient autonomy, and affirm the value and ability of healthy lifestyle behavior will likely negate the adverse impacts of stigma on health and well-being.

The WHISTLE model extends behavioral change frameworks. For example, motivational interviewing and self-determination theory (i.e., an empirical framework for enhancing behavior change) are conceptually related and aid practitioners in helping patients internalize their own motivation and ultimately foster volitional health behavior change (Abildsnes et al., 2021). Although these informed the development of the WHISTLE model, this model uniquely highlights the necessity of averting weight stigma's disrupting role in hampering practitioner-patient communication and provides a feasible template for accomplishing this task. It is also important to recognize that increasing sensitivity to perceived harm, such as weight stigma, may engender a sense of diminished perceived control among patients, which is a factor that has long been recognized as a facilitator of health behavior change

(Schwarzer & Fuchs, 1996; Strecher et al., 1986). Conversely, while self-affirmation can be a barrier, it is also a catalyst for behavior change (Cohen & Sherman, 2014). Indeed, research has demonstrated the value of affirmation interventions, which are associated with better health, including increased health behavior and even lowering body weight toward healthier ranges (Epton et al., 2015; Ferrer & Cohen, 2019; Logel & Cohen, 2012). For example, self-affirmation is associated with greater acceptance of health information and stronger motivation for health behavior change when receiving health risk information (Epton et al., 2015).

Discussion

The WHISTLE model is not a full-scale intervention but rather emphasizes an attitude of acceptance in which practitioners harmoniously attend to disarming stigma and facilitating health behavior for weight-related health. Other researchers have developed a high-quality guide to foster practitioner-patient communication within a motivational interviewing approach to reduce stigma in weight health (see Scherr et al., 2023, for a review). The WHISTLE model adds to this research by focusing on attitudes and actions of acceptance, and its primary purpose is to initiate conversations about weight health and behavioral change within an attitude of acceptance to prevent and reduce stigma. Beyond initial engagement and discussion about weight health using the WHISTLE model, practitioners use methods and treatments congruent with their expertise, responsibility, and setting. More simply, a broad scope guideline to continue conversations is outlined within a Brief Action Plan (BAP), which is a person-centered approach to facilitate health behavioral change and has supporting evidence across many healthcare settings (Jadotte et al., 2023). The WHISTLE model's biopsychosocial lens allows for a compassionate and health-focused approach to reduce stigma and empower patients to participate in healthy lifestyle behaviors. Consistent with motivational interviewing principles, the WHISTLE model does not imply endorsement of unhealthy behaviors but rather acts as a relational process in which an attitude of acceptance deactivates stigma while still facilitating health behavior change. As practitioners within the WHISTLE model focus on empirically supported health behaviors in an atmosphere of acceptance that affirms one's inherent value, defuses stigma, and supports autonomy for health behavior change, weight-related health is likely to improve.

A Call to Healthcare Practitioners

Psychological theories on behavior change are advancing the landscape of healthcare quality (Hilton, 2023). Healthcare practitioners across training backgrounds and treatment settings have a duty to promote weight health and well-being. Mental health clinicians are even emerging as important, supplementary practitioners to weight-related health concerns (Dandgey & Patten, 2023; Murray et al., 2021). Given the rising adverse health consequences of higher body weight, inaction will have negative global impacts on health, well-being, and economic resources (Okunogbe et al., 2022). Although calls for systemic-based changes are increasing (and likely required) to effectuate lasting weight-related health improvements from a public health perspective (Okunogbe et al., 2022; Puhl & Suh, 2015), practitioners maintain the responsibility to promote health behavior change.

Health psychology research has recognized weight-related health as a persistent public health concern despite decades of targeted resources and research (Brownell, 2010). Health psychologists have a duty to bilaterally support patients based on the advancing empirical literature across health sciences. Just as weight stigma research is progressing, the data establishing the health benefits of healthy lifestyle behavior and the risks for morbidity and mortality related to higher body weight are also advancing. Health psychologists are well-positioned to handle potentially conflicting goals related to stigma, behavior, and weight health. However, to maintain this position, health psychologists and related professionals must maintain fidelity to evidence demonstrating the negative impacts of higher body weight and strive to mitigate risks through facilitating health behavior. For example, contrary to HAES, health psychologists cannot maintain weight-related neutrality when presented with opportunities to promote weight-related health. Health behavior and weight-related health dialogue need not be removed nor disparaged to reduce weight-related stigma. Promoting empirically supported health behaviors (e.g., physical activity) for weight-related health is vital for patients and a duty of practitioners.

Future Research Directions

The WHISTLE model has implications for at least three directions for research. First, there has been promising research on self-affirmation as a weight-related health treatment (Logel & Cohen, 2012). More research should test whether self-affirmation interventions are effective for protecting patient self-identity in the context of receiving healthcare recommendations, reducing weight stigma, and facilitating health behavior changes congruent with weight health recommendations. Relatedly, compounding stigmas represent a greater health concern (Stangl et al., 2019), but research has only recently begun to investigate weight stigma and diversity, such as the intersectionality of weight stigma, race, and gender. Experienced weight stigma is similar across racial and gender groups but internalized less among Black and Hispanic individuals relative to White individuals, and less among men relative to women (Himmelstein et al., 2017; Reece, 2019; Wetzel & Himmelstein, 2024). Notably, lower self-compassion is associated with greater weight stigma across racial and socioeconomic backgrounds, which represents a potential target for reducing internalized weight stigma (Puhl et al., 2020).

Second, the impact of technical aspects of the WHISTLE model (i.e., acceptance statements and questions) should be empirically investigated to test the hypothesized effect of simultaneously reducing weight stigma and facilitating healthy lifestyle behavioral change (Moizé et al., 2025). To operationalize the processes of the WHISTLE model, researchers could determine whether a practitioner's statement (or question) for initiating conversations about weight health is congruent with the four 'A's of acceptance. For example, 1 = congruent, 0 = partially congruent, and -1 = not congruent, with higher total scores reflecting higher adherence to the WHISTLE model. Table 1 provides appropriate statements, each of which is congruent with assumptions, attitudes, and actions of acceptance in the WHISTLE model. A randomized controlled trial could be conducted to compare the weight stigma and behavioral outcomes of practitioners who use the WHISTLE model to initiate conversations with patients compared to an active control, such as practitioners' treatment as usual, which often involves blunt recommendations to reduce body weight that exacerbate stigma (Standen et al., 2024). Behavioral outcomes may be measured by any relevant health behavior metric (e.g., change in total minutes per week spent participating in physical activity). Weight stigma outcomes may be assessed using existing validated weight stigma measures such as the Perceived Weight Stigma Scale (PWSS; Schafer & Ferraro, 2011).

Third, researchers from disparate training backgrounds (e.g., medicine, public health, psychology, sociology) may consider joint initiatives to synthesize data across biopsychosocial health science domains to elucidate weight health and stigma interactions. Results of such research would inform interventions to simultaneously reduce stigma and promote weight health and well-being.

A major limitation is that extant literature on motivational interviewing has not examined its potential effectiveness (i.e., its potential causal role) for reducing weight stigma, as evidenced by a systematic rapid review (see Appendix) using Approach 3 by Tricco et al. (2016). However, there are theoretical reasons that suggest that acceptance would be potentially effective for reducing stigma given that motivational interviewing is efficacious for behavioral change among people with substance use issues, who experience substance use stigma (El Hayek et al., 2024; Kulesza et al., 2013; Magnan et al., 2024; Miller & Rollnick, 2023; Yang et al., 2018). The only study to quantitatively correlate motivational interviewing and weight stigma showed that physician experience with motivational interviewing was associated with more use of person-first language and positive perceptions of patient adherence to treatment recommendations (Bennett & Puhl, 2024). Although the WHISTLE model's foundation on patient acceptance and engagement lends itself well to reducing weight stigma in healthcare contexts (Moizé et al., 2025), the empirical research previously outlined is a necessary next step to evaluate the potential efficacy of the WHISTLE model.

Conclusion

The WHISTLE model is utilitarian, emphasizing action where health behavior promotion and an attitude of acceptance to reduce stigma occur concurrently. At its basic level, the WHISTLE model is a unifying agent for practitioners and patients working together to improve weight-related health. Practitioners must engage in accepting and collaborative conversations to increase health behavior, whether preventive or reactive, focusing on person-centered care to foster weight health. Attitudes and actions demonstrated in an atmosphere of acceptance are likely to lead to reduced stigma and enhanced practitioner-patient communication. Healthcare and mental health practitioners who encourage health behavior and discuss weight health may face criticism from colleagues or patients who favor reducing stigma over promoting weight-related health as campaigns against weight stigma become increasingly heard and promoted (e.g., HAES). However, there is no need to campaign for avoiding conversations about health behavior and denigrate health science research for the destigmatization of weight. Indeed, discussing the importance of fundamental health behavior (e.g., physical activity) may appear less popular, but it is more vital than ever. Overall, the WHISTLE model functions to increase patient-centered care by empowering practitioners to face the challenge of initiating weight-related health and health behavior change discussions in a way that deactivates the barriers of stigma.

Acknowledgements

I sincerely thank Dr. Andrew H. Hales for his professorship and writing mentorship.

References

Abildsnes, E., Elin Andresen, N., Storbaekken, S., Beate Samdal, G., Mildestvedt, T., & Meland, E. (2021). How is motivational interviewing (un) related to self-determination theory: An empirical study from different healthcare settings. *Scandinavian Journal of Psychology, 62*(5), 709–716. https://doi.org /10.1111/sjop.12757

Allison, D. B., Downey, M., Atkinson, R. L., Billington, C. J., Bray, G. A., Eckel, R. H., Finkelstein, E. A., Jensen, M. D., & Tremblay, A. (2008). Obesity as a disease: A white paper on evidence and arguments commissioned by the Council of the Obesity Society. *Obesity, 16*(6), 1161–1177. https://doi.org/10.1038/oby.2008.231

Amiri, P., Mansouri-Tehrani, M. M., Khalili-Chelik, A., Karimi, M., Jalali-Farahani, S., Amouzegar, A., & Kazemian, E. (2022). Does motivational interviewing improve the weight management process in adolescents? A systematic review and meta-analysis. *International Journal of Behavioral Medicine, 29*(1), 78–103. https://doi.org/10.1007/s12529-021-09994-w

Armstrong, D. (2014). Actors, patients and agency: A recent history. *Sociology of Health & Illness, 36*(2), 163–174. https://doi.org/10.1111/1467-9566.12100

Bennett, B. L., & Puhl, R. M. (2024). Physicians' stigmatizing attitudes about individuals with type 2 diabetes: Associations with communication practices and perceived barriers to care. *Primary Care Diabetes, 18*(5), 518–524. https://doi.org/10.1016/j.pcd.2024.07.001

Bischof, G., Bischof, A., & Rumpf, H. J. (2021). Motivational Interviewing: An evidence-based approach for use in medical practice. *Deutsches Arzteblatt International, 118*(7), 109–115. https://doi.org/10.3238/arztebl.m2021.0014

Bombak, A. (2014). Obesity, health at every size, and public health policy. *American Journal of Public Health, 104*(2), e60–e67. https://doi.org/10.2105/AJPH.2013.301486

Brownell, K. D. (2010). The humbling experience of treating obesity: Should we persist or desist?. *Behaviour Research and Therapy, 48*(8), 717–719. https://doi.org/10.1016/j.brat.2010.05.018

Callahan, D. (2013). Obesity: Chasing an elusive epidemic. *The Hastings Center Report, 43*(1), 34–40. https://doi.org/10.1002/hast.114

Cohen, G. L., & Sherman, D. K. (2014). The psychology of change: Self-affirmation and social psychological intervention. *Annual Review of Psychology, 65*, 333–371. https://doi.org/10.1146/annurev-psych-010213-115137

Copeland, L., McNamara, R., Kelson, M., & Simpson, S. (2017). The development of a coding system to code planning talk within motivational interviewing. *Patient Education and Counseling, 100*(2), 313–319. https://doi.org/10.1016/j.pec.2016.09.003

Dandgey, S., & Patten, E. (2023). Psychological considerations for the holistic management of obesity. *Clinical Medicine, 23*(4), 318–322. https://doi.org/10.7861/clinmed.2023-0146

Elgaddal, N., Kramarow, E. A., & Reuben, C. (2022). Physical activity among adults aged 18 and over: United States, 2020. *National Center for Health Statistics*. NCHS Data Brief, no 443. https://dx.doi.org/10.15620/cdc:120213

El Hayek, S., Foad, W., de Filippis, R., Ghosh, A., Koukach, N., Mahgoub Mohammed Khier, A., Pant, S. B., Padilla, V., Ramalho, R., Tolba, H., & Shalbafan, M. (2024). Stigma toward substance use disorders: A multinational perspective and call for action. *Frontiers in Psychiatry, 15*, Article 1295818. https://doi.org/10.3389/fpsyt.2024.1295818

Epton, T., Harris, P. R., Kane, R., van Koningsbruggen, G. M., & Sheeran, P. (2015). The impact of self-affirmation on health-behavior change: A meta-analysis. *Health Psychology, 34*(3), 187–196. https://doi.org/10.1037/hea0000116

Ferrer, R. A., & Cohen, G. L. (2019). Reconceptualizing self-affirmation with the trigger and channel framework: Lessons from the health domain. *Personality and Social Psychology Review, 23*(3), 285-304. https://doi.org/10.1177/1088868318797036

Frost, H., Campbell, P., Maxwell, M., O'Carroll, R. E., Dombrowski, S. U., Williams, B., Cheyne, H., Coles, E., & Pollock, A. (2018). Effectiveness of motivational interviewing on adult behaviour change in health and social care settings: A systematic review of reviews. *PloS One, 13*(10), Article e0204890. https://doi.org/10.1371/journal.pone.0204890

Fruh, S. M. (2017). Obesity: Risk factors, complications, and strategies for sustainable long-term weight management. *Journal of the American Association of Nurse Practitioners, 29*(S1), S3–S14. https://doi.org/10.1002/2327-6924.12510

Haase, C. L., Lopes, S., Olsen, A. H., Satylganova, A., Schnecke, V., & McEwan, P. (2021). Weight loss and risk reduction of obesity-related outcomes in 0.5 million people: Evidence from a UK primary care database. *International Journal of Obesity, 45*(6), 1249–1258. https://doi.org/10.1038/s41366-021-00788-4

Hatzenbuehler, M. L., Phelan, J. C., & Link, B. G. (2013). Stigma as a fundamental cause of population health inequalities. *American Journal of Public Health, 103*(5), 813–821. https://doi.org/10.2105/AJPH.2012.301069

Himmelstein, M. S., Puhl, R. M., & Quinn, D. M. (2017). Intersectionality: An Understudied Framework for Addressing Weight Stigma. *American Journal of Preventive Medicine, 53*(4), 421–431. https://doi.org/10.1016/j.amepre.2017.04.003

Hruby, A., & Hu, F. B. (2015). The epidemiology of obesity: A big picture. *PharmacoEconomics, 33*(7), 673–689. https://doi.org/10.1007/s40273-014-0243-x

Jadotte, Y., Buchholz, B., Carroll, W., Frum-Vassallo, D., MacPherson, J., & Cole, S. (2023). Brief action planning in health and health care: A scoping review. *The Medical Clinics of North America, 107*(6), 1047–1096. https://doi.org/10.1016/j.mcna.2023.06.018

Katz, D. L. (2014). Perspective: Obesity is not a disease. *Nature, 508*(7496), S57. https://doi.org/10.1038/508S57a

Kulesza, M., Larimer, M. E., & Rao, D. (2013). Substance use related stigma: What we know and the way forward. *Journal of Addictive Behaviors, Therapy & Rehabilitation, 2*(2), 782. https://doi.org/10.4172/2324-9005.1000106

Logel, C., & Cohen, G. L. (2012). The role of the self in physical health: Testing the effect of a values-affirmation intervention on weight loss. *Psychological Science, 23*(1), 53–55. https://doi.org/10.1177/0956797611421936

Lundahl, B., Moleni, T., Burke, B. L., Butters, R., Tollefson, D., Butler, C., & Rollnick, S. (2013). Motivational interviewing in medical care settings: A systematic review and meta-analysis of randomized controlled trials. *Patient Education and Counseling, 93*(2), 157–168. https://doi.org/10.1016/j.pec.2013.07.012

Luty, J., & Iwanowicz, M. (2018). Motivational interviewing: Living up to its promise? *BJPsych Advances, 24*(1), 46–53. https://doi.org/10.1192/bja.2017.7

Magill, M., Apodaca, T. R., Borsari, B., Gaume, J., Hoadley, A., Gordon, R. E. F., Tonigan, J. S., & Moyers, T. (2018). A meta-analysis of motivational interviewing process: Technical, relational, and conditional process models of change. *Journal of Consulting and Clinical Psychology, 86*(2), 140–157. https://doi.org/10.1037/ccp0000250

Magnan, E., Weyrich, M., Miller, M., Melnikow, J., Moulin, A., Servis, M., Chadha, P., Spivack, S., & Henry, S. G. (2024). Stigma against patients with substance use disorders among health care professionals and trainees and stigma-reducing interventions: A systematic review. *Academic Medicine, 99*(2), 221–231. https://doi.org/10.1097/ACM.0000000000005467

Major, B., Hunger, J. M., Bunyan, D. P., & Miller, C. T. (2014). The ironic effects of weight stigma.

Journal of Experimental Social Psychology, 51, 74–80. https://doi.org/10.1016/j.jesp.2013.11.009

Makin, H., Chisholm, A., Fallon, V., & Goodwin, L. (2021). Use of motivational interviewing in behavioural interventions among adults with obesity: A systematic review and meta-analysis. *Clinical Obesity, 11*(4), Article e12457. https://doi.org/10.1111/cob.12457

Marks, D. F. (2015). Homeostatic theory of obesity. *Health Psychology Open, 2*(1). https://doi.org/10.1177/2055102915590692

McCabe, M., Alcaraz-Ibanez, M., Markey, C., Sicilia, A., Rodgers, R. F., Aimé, A., ... Fuller-Tyszkiewicz, M. (2023). A longitudinal evaluation of a biopsychosocial model predicting BMI and disordered eating among young adults. *Australian Psychologist, 58*(2), 57–79. https://doi.org/10.1080/00050067.2023.2181686

Michalopoulou, M., Ferrey, A. E., Harmer, G., Goddard, L., Kebbe, M., Theodoulou, A., Jebb, S. A., & Aveyard, P. (2022). Effectiveness of motivational interviewing in managing overweight and obesity: A systematic review and meta-analysis. *Annals of Internal Medicine, 175*(6), 838–850. https://doi.org/10.7326/M21-3128

Michie, S., Atkins, L. & West, R. (2014). *The Behaviour Change Wheel: A guide to designing interventions*. Silverback Publishing.

Miller, W. R., & Rollnick, S. (2013). *Motivational interviewing: Helping people change* (3rd ed.). Guilford Press.

Miller, W. R., & Rollnick, S. (2023). *Motivational interviewing: Helping people change and grow* (4th ed.). Guilford Press.

Murray, K., Brichacek, E., & Rieger, E. (2021). Intentions to see a psychologist for weight management: A cross-sectional study investigating beliefs about psychologists, psychological services and obesity. *Journal of Clinical Psychology, 77*(10), 2288–2305. https://doi.org/10.1002/jclp.23163

Moizé, V., Graham, Y., Ramos Salas, X., & Balcells, M. (2025). Motivational interviewing (MI) in obesity care: Cultivating person-centered and supportive clinical conversations to reduce stigma: A narrative review. *Obesity Science & Practice, 11*(1). https://doi.org/10.1002/osp4.678

Okunogbe, A., Nugent, R., Spencer, G., Powis, J., Ralston, J., & Wilding, J. (2022). Economic impacts of overweight and obesity: Current and future estimates for 161 countries. *BMJ Global Health, 7*(9), Article e009773. https://doi.org/10.1136/bmjgh-2022-009773

Ortega, F. B., Lee, D. C., Katzmarzyk, P. T., Ruiz, J. R., Sui, X., Church, T. S., & Blair, S. N. (2013). The intriguing metabolically healthy but obese phenotype: Cardiovascular prognosis and role of fitness. *European Heart Journal, 34*(5), 389–397. https://doi.org/10.1093/eurheartj/ehs174

Panza, G. A., Puhl, R. M., Taylor, B. A., Cilhoroz, B., Himmelstein, M. S., Fernandez, A. B., & Pescatello, L. S. (2023). The effects of an acute weight stigma exposure on cardiovascular reactivity among women with obesity and hypertension: A randomized trial. *Journal of Psychosomatic Research, 165*, Article 111124. https://doi.org/10.1016/j.jpsychores.2022.111124

Penney, T. L., & Kirk, S. F. (2015). The Health at Every Size paradigm and obesity: Missing empirical evidence may help push the reframing obesity debate forward. *American Journal of Public Health, 105*(5), e38–e42. https://doi.org/10.2105/AJPH.2015.302552

Pi-Sunyer, X. (2009). The medical risks of obesity. *Postgraduate Medicine, 121*(6), 21–33. https://doi.org/10.3810/pgm.2009.11.2074

Plotnikoff, R. C., Bercovitz, K., Rhodes, R. E., Loucaides, C. A., & Karunamuni, N. (2007). Testing a conceptual model related to weight perceptions, physical activity and smoking in adolescents. *Health Education Research, 22*(2), 192–202. https://doi.org/10.1093/her/cyl065

Pojednic, R., D'Arpino, E., Halliday, I., & Bantham, A. (2022). The benefits of physical activity for people with obesity, independent of weight loss: A systematic review. *International Journal of Environmental Research and Public Health, 19*(9), Article 4981. https://doi.org/10.3390/ijerph19094981

Pont, S. J., Puhl, R., Cook, S. R., Slusser, W., Section on Obesity, & Obesity Society (2017). Stigma experienced by children and adolescents with obesity. *Pediatrics, 140*(6), Article e20173034. https://doi.org /10.1542/peds.2017-3034

Puhl R. M. (2023). Weight stigma and barriers to effective obesity care. *Gastroenterology clinics of North America, 52*(2), 417–428. https://doi.org/10.1016/j.gtc.2023.02.002

Puhl, R., & Brownell, K. D. (2001). Bias, discrimination, and obesity. *Obesity Research, 9*(12), 788–805. https://doi.org/10.1038/oby.2001.108

Puhl, R. M., Lessard, L. M., Himmelstein, M. S., & Foster, G. D. (2021). The roles of experienced and internalized weight stigma in healthcare experiences: Perspectives of adults engaged in weight management across six countries. *PloS One, 16*(6), Article e0251566. https://doi.org/10.1371/journal.pone.0251566

Puhl, R., & Suh, Y. (2015). Health consequences of weight stigma: Implications for obesity prevention and treatment. *Current Obesity Reports, 4*(2), 182–190. https://doi.org/10.1007/s13679-015-0153-z

Puhl, R. M., Telke, S., Larson, N., Eisenberg, M. E., & Neumark-Stzainer, D. (2020). Experiences of weight stigma and links with self-compassion among a population-based sample of young adults from diverse ethnic/racial and socio-economic backgrounds. *Journal of Psychosomatic Research, 134*, Article 110134. https://doi.org/10.1016/j.jpsychores.2020.110134

Reece, R. L. (2019). Coloring weight stigma: On race, colorism, weight stigma, and the failure of additive intersectionality. *Sociology of Race and Ethnicity, 5*(3), 388-400. https://doi.org/10.1177/2332649218795185

Robinson, E., Roberts, C., Vainik, U., & Jones, A. (2020). The psychology of obesity: An umbrella review and evidence-based map of the psychological correlates of heavier body weight. *Neuroscience and Biobehavioral Reviews, 119*, 468–480. https://doi.org/10.1016/j.neubiorev.2020.10.009

Rubak, S., Sandbaek, A., Lauritzen, T., & Christensen, B. (2005). Motivational interviewing: A systematic review and meta-analysis. *The British Journal of General Practice, 55*(513), 305–312. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1463134/

Rubino, F., Puhl, R. M., Cummings, D. E., Eckel, R. H., Ryan, D. H., Mechanick, J. I., Nadglowski, J., Ramos Salas, X., Schauer, P. R., Twenefour, D., Apovian, C. M., Aronne, L. J., Batterham, R. L., Berthoud, H. R., Boza, C., Busetto, L., Dicker, D., De Groot, M., Eisenberg, D., Flint, S. W., ... Dixon, J. B. (2020). Joint international consensus statement for ending stigma of obesity. *Nature Medicine, 26*(4), 485–497. https://doi.org/10.1038/s41591-020-0803-x

Ryan, D. H., & Yockey, S. R. (2017). Weight loss and improvement in comorbidity: Differences at 5%, 10%, 15%, and over. *Current Obesity Reports, 6*(2), 187–194. https://doi.org/10.1007/s13679-017-0262-y

Safaei, M., Sundararajan, E. A., Driss, M., Boulila, W., & Shapi'i, A. (2021). A systematic literature review on obesity: Understanding the causes & consequences of obesity and reviewing various machine learning approaches used to predict obesity. *Computers in Biology and Medicine, 136*, Article 104754. https://doi.org/10.1016/j.compbiomed.2021.104754

Scarff, J. R. (2021). What's in a name? The problematic term "Provider". *Federal Practitioner, 38*(10), 446–448. https://doi.org/10.12788/fp.0188

Schafer, M. H., & Ferraro, K. F. (2011). The stigma of obesity: Does perceived weight discrimination affect identity and physical health? *Social Psychology Quarterly, 74*(1), 76-97. https://doi.org/10.1177/0190272511398197

Scherr, K. A., Honeycutt, L. K., Page, S., & Armstrong, S. C. (2023). Communication, bias, and stigma. In C. K. Fox (Ed.), *Managing pediatric obesity using advanced therapies: Practical guide for pediatric health care providers* (pp. 27–51). Springer International Publishing.

Schumacher, J. A., & Madson A. B. (2014). *Fundamentals of motivational interviewing: Tips and strategies for addressing common clinical challenges*. Oxford University Press.

Schwarzer, R., & Fuchs, R. (1996). Self-efficacy and health behaviours. In M. Conner & P. Norman (Eds.), *Predicting health behaviour: Research and practice with social cognition models* (pp. 163–196). Open University Press.

Standen, E. C., Rothman, A. J., & Mann, T. (2024). Consequences of receiving weight-related advice from a healthcare provider: Understanding the varied experiences of people with higher weight. *Social Science & Medicine, 347*, Article 116784. https://doi.org/10.1016/j.socscimed.2024.116784

Strecher, V. J., DeVellis, B. M., Becker, M. H., & Rosenstock, I. M. (1986). The role of self-efficacy in achieving health behavior change. *Health Education Quarterly, 13*(1), 73–92. https://doi.org/10.1177/109019818601300108

Tomiyama, A. J. (2014). Weight stigma is stressful. A review of evidence for the Cyclic Obesity/Weight-Based Stigma model. *Appetite, 82*, 8–15. https://doi.org/10.1016/j.appet.2014.06.108

Tricco, A. C., Zarin, W., Antony, J., Hutton, B., Moher, D., Sherifali, D., & Straus, S. E. (2016). An international survey and modified Delphi approach revealed numerous rapid review methods. *Journal of Clinical Epidemiology, 70*, 61–67. https://doi.org/10.1016/j.jclinepi.2015.08.012

Visscher, T. L., & Seidell, J. C. (2001). The public health impact of obesity. *Annual Review of Public Health, 22*, 355–375. https://doi.org/10.1146/annurev.publhealth.22.1.355

Wetzel, K. E., & Himmelstein, M. S. (2024). Weight stigma is uniquely tied to maladaptive eating across different racial, ethnic, and gender groups. *Appetite, 201*, Article 107604. https://doi.org/10.1016/j.appet.2024.107604

Yang, L. H., Wong, L. Y., Grivel, M. M., & Hasin, D. S. (2017). Stigma and substance use disorders: An international phenomenon. *Current Opinion in Psychiatry, 30*(5), 378–388. https://doi.org/10.1097/YCO.0000000000000351

Figure 1.

The Weight Health through Integrated Stigma-Reduction and Lifestyle Engagement (WHISTLE) Model

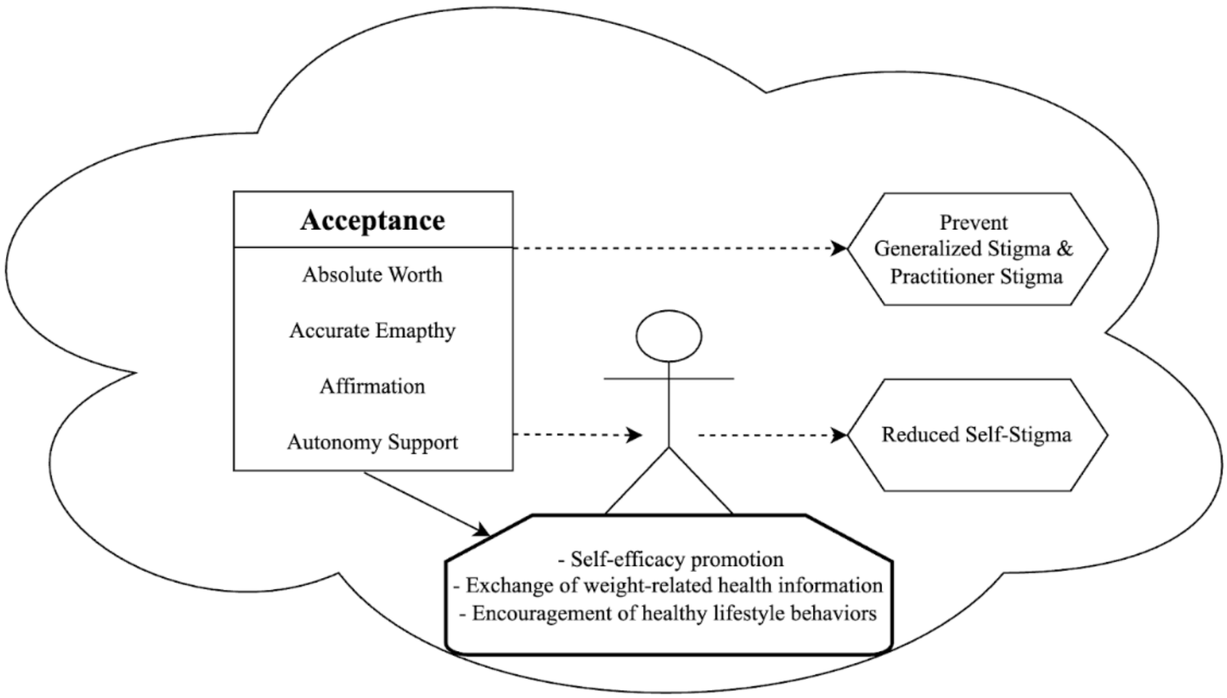


Table 1.

Sample Statements to Initiate Weight Health Conversations Congruent with Acceptance

Absolute Worth
- I care about you and your health regardless of your current body weight or what your physical activity and dietary habits look like right now. If you'd be willing, could we talk about some things that research suggests might support weight health in ways that are meaningful to you? - Your importance as a person does not depend on how much you weigh, how often you go for a walk, or what you eat. Your health, however, can be linked to what you do. If you're up to it, I'd like to hear your thoughts on how you feel about your weight, health, and lifestyle health behaviors right now.
Accurate Empathy
- I can see how it would be frustrating for you to be productive at work when you're feeling judged for your body size and what you eat. And if it would be of interest to you, could we talk together about research on weight and health to see what may be helpful and relevant for you? - It sounds hard to feel excluded from activities like softball that you used to enjoy with friends. I'd like to hear more about that, and if you'd be open to it, we could also explore other activities you might find enjoyable, and that could support your weight health.
Affirmation
- You've been working hard to manage your health condition, and that's really impressive! If it would be okay with you, we could talk about how body weight and physical activity can further support the progress you're making. - Coming in today (e.g., insert any healthcare-related appointment) shows how much you care about your health. Would it be okay if we talked about where weight health fits into what's most important to you?
Autonomy Support
- Over 99% of your life is spent outside this clinic (e.g., insert setting), and I imagine you have some great ideas about what might work best for your health. What are your thoughts on how your health goals, including weight, fit into the bigger picture for you? - I realize that it can be hard to talk about weight health. I want you to know that I'm here to support you as you consider making lifestyle changes that *you* think would be helpful. - You clearly put effort into taking care of your health, which is great. If it sounds okay to you, we could explore how your weight may play a role in what's going on with your health right now and what steps *you* might consider for your health.

Note. Attuned readers may notice the conceptual overlap of the *acceptance* statements and even their near interchangeability. After each of these statements, an additional acknowledgement of patient autonomy may fit as well. For example, "And, of course, any changes (e.g., related to health, physical activity, eating) you choose to make would be completely up to you."

Appendix

Supporting Table 1. *Search strategy for PubMed*

Block 1: *Motivational interviewing*

Search name	Search query	Type of search
1	("motivational interviewing"[MeSH Terms] OR "motivation* interview*"[Title/Abstract] OR "mi style"[Title/Abstract] OR "motivation* intervention*"[Title/Abstract] OR "motivational counseling"[Title/Abstract] OR "motivational counselling"[Title/Abstract])	MeSH terms and Title/Abstract

Block 2: *Stigma*

Search name	Search query	Type of search
2	("social stigma"[MeSH Terms] OR "weight prejudice"[MeSH Terms] OR "social discrimination"[MeSH Terms] OR "perceived discrimination"[MeSH Terms] OR "fat sham*"[Title/Abstract] OR "fat sham*"[Title/Abstract] OR "weight stigma*"[Title/Abstract] OR "weight stigma*"[Title/Abstract] OR "fat stigma*"[Title/Abstract] OR "fat stigma*"[Title/Abstract])	MeSH terms and Title/Abstract

Block 3: *Weight*

Search name	Search query	Type of search
3	("obesity"[MeSH Terms] OR "overweight"[MeSH Terms] OR "body weight"[MeSH Terms] OR "weight loss"[MeSH Terms] OR "ideal body weight"[MeSH Terms] OR "weight health"[Title/Abstract])	MeSH terms and Title/Abstract

Note. Database was accessed 4/2/2025. No filters or date ranges were applied. The three blocks were connected with the Boolean operator 'AND'. The search yielded five results. Screening determined that no study examined the potential relationship between motivational interviewing and weight stigma.

Appendix

Supporting Table 2. *Search strategy for Scopus*

Block 1: *Motivational interviewing*

Search name	Search query	Type of search
1	("motivational interviewing" OR "motivation* interview*" OR "mi style" OR "motivation* intervention*" OR "motivational counseling" OR "motivational counselling")	Title, abstract, keyword

Block 2: *Stigma*

Search name	Search query	Type of search
2	("social stigma" OR "weight prejudice" OR "social discrimination" OR "perceived discrimination" OR "fat sham*" OR "fat sham*" OR "weight stigma*" OR "weight stigma*" OR "fat stigma*" OR "fat stigma*")	Title, abstract, keyword

Block 3: *Weight*

Search name	Search query	Type of search
3	("obesity" OR "overweight" OR "body weight" OR "weight loss" OR "ideal body weight" OR "weight health"])	Title, abstract, keyword

Note. Database was accessed 4/2/2025. No filters or date ranges were applied. The three blocks were connected with the Boolean operator 'AND'. The search yielded five results. Screening determined that no study examined the potential relationship between motivational interviewing and weight stigma.

The Pertinent Prenatal Period: A Secondary Analysis Examining the Relationship Between Prenatal Maternal Anxiety and Child Language Development at 18 Months

Melanie Lucia Esterine
Department of Counseling and Clinical Psychology, Teachers College at Columbia University, New York, NY USA

Maternal mental health is thought to be an important factor that may shape child development as early as the prenatal period. Prenatal maternal mental health has been linked to both physiological and factors that are theorized to explain the links between mental health and child language development. Whereas some past research has examined the associations between prenatal mental health and child language development, most have examined broad measures that include both anxiety and depression. Given that there are distinct mechanisms by which these aspects of maternal mental health are hypothesized to impact development, this study aims to parse the association between anxiety and child language development specifically. This study utilized data from a longitudinal prospective study to examine the associations between prenatal anxiety and child language development at 18 months in 167 families. We conducted a series of three regression models beginning with a simple linear regression examining prenatal anxiety and language development, followed by two models, first adding demographic covariates, then including prenatal depression. We did not find associations between maternal anxiety and child language development; inadvertently raising the question of when and by which mechanisms maternal mental health may or may not impact aspects of child development, which are crucial answers to be discerned in order to determine the most effective way to support pregnant mothers and their children throughout the prenatal period.

Keywords: prenatal anxiety, language development, maternal mental health, child outcomes

Prenatal anxiety, an issue impacting pregnant persons with a prevalence ranging from 21% to 25% (Field, 2017), could potentially play an important role in shaping child language development. While few studies have examined the links between anxiety and language development, existing studies suggest that greater anxiety is associated with poorer development (Ibanez et al., 2015). In studies that have examined prenatal anxiety as an independent variable, maternal mental health tends to be discussed broadly without consideration for distinct aspects of prenatal mental health (e.g., anxiety versus depression; Rogers et al., 2020). Moreover, there has been a focus on child cognitive outcomes such as memory or learning rather than language development (Keim et al., 2011).

One prominent study exploring the association between prenatal mental health, as encompassed by depression and anxiety, and children's cognitive outcomes found that prenatal anxiety was significantly associated with lower scores on assessments of cognition (Ibanez et al., 2015). In this study, the MacArthur Communicative Development Inventory (CDI) was used to assess language as a cognitive domain, and it was observed that higher levels of prenatal anxiety were associated with lower CDI scores in children (Ibanez et al., 2015). However, this study did not disentangle the influence of anxiety from that of depression. Additionally, the CDI is based entirely on maternal reports, which may introduce outcomes such as social desirability bias or the inability to accurately attribute differences in reported vocabulary skills to true differences in skill, in addition to the possibility that more anxious mothers may simply report differently than non-anxious mothers.

Additionally, two studies investigated the trajectory of maternal prenatal anxiety throughout pregnancy and developmental outcomes in 1-year-old infants (Irwin et al., 2020; Keim et al., 2011). These studies examined prenatal anxiety as an independent variable but did not control for varying mental health disorders that are highly comorbid. Both studies found that higher prenatal anxiety was associated with lower scores on assessments of cognition. Specifically, Irwin et al. (2020) found that increasing maternal anxiety across pregnancy was associated with lower receptive language, also known as the "input" of language.

Next, a study conducted by Rogers et al. (2020) aimed to assess whether maternal prenatal depression and anxiety were adversely associated with varying developmental skills in children during the first 18 years of life. They found that prenatal anxiety was associated with poorer language abilities in early childhood, as measured through a composite, without controls for varying mental health disorders.

In a study that controlled for confounding mental health comorbidities, Brouwers et al. (2001) showed that high prenatal anxiety during late pregnancy was uniquely associated with lower attention-related process skills at three weeks, one year, and two years. This finding may lend insight to investigations of language acquisition, given that attention-related processes are thought to be related to language acquisition and processing abilities (de Diego-Balaguer et al., 2016; Kannass & Oakes, 2008). Finally, a pertinent study conducted by Clifford et al. (2022), which examined the association between postnatal maternal mental health and children's expressive language (i.e., output of language) at approximately one and a half years of age (Mage = 17.03 months), found that postnatal maternal anxiety was negatively related to child language production scores, as reported by a parent. In this study, prenatal anxiety is not the independent variable; however, its relevance lies in the dependent variables falling within the domain of language.

Researchers have theorized that both physiological and psychological processes could underlie the associations between prenatal anxiety and postnatal child development. Physiological issues have been proposed as one mechanism by which prenatal anxiety could shape development. Notably, greater prenatal anxiety has been associated with poorer fetal development, including reduced fetal head circumference (Lewis et al., 2015), with slower or decreased fetal head growth during gestation associated with language difficulties in children two years of age compared to their typically developed peers (Villar et al., 2021). Research has also found that extremely high levels of prenatal anxiety may restrict blood flow and oxygen to developing fetal organs (Hobel & Culhane, 2003). Restricted blood flow in particular can cause Maternal Vascular Malperfusion, a placental abnormality, which has been linked to poor language development (Straughen et al., 2017). These findings highlight physiological pathways through which prenatal anxiety may implicate a child's language abilities.

It is also possible that prenatal anxiety forecasts or precedes less developmentally supportive parenting behaviors. Research has demonstrated prenatal anxiety to be a distinct predictor of postnatal mood disturbance (Blackmore et al., 2016), which has the potential to negatively influence child development, including language development, through the possible presence of these less developmentally supportive parenting behaviors. The Family Stress Model, originally introduced by Conger et al. (1992), suggests that parental psychological distress, including anxiety, can lead to parenting practices that may negatively impact child development (Masarik & Conger, 2017). Developmentally unsupportive parenting practices may include insensitivity (Newland et al., 2013); reduced time and quality of time spent with children (Iruka et al., 2012); and harsh, punitive, and over-controlling behaviors (Emmen et al., 2013) that increase the risk of child abuse or neglect (Warren & Font, 2015). Furthermore, maladaptive maternal behaviors and affect, as encompassed by diminished mental well-being, yield a less stimulating environment for children as it relates to the home language environment (i.e., the language that happens in the home, including the words that the child hears and reciprocal back-and-forth parent-child conversations; Clifford et al., 2021). Research has suggested that the quality of a child's early home language environment is a crucial predictor of their later linguistic skills and cognitive development (Dailey & Bergelson, 2022; Romeo et al., 2018). It is the exposure to more words and engagement that yield more successful linguistic developmental outcomes, such as the ability to learn vocabulary faster, exhibit increased processing speed, and produce overall stronger language (Gilkerson & Richards, 2009; Hart et al., 1997; Hurtado et al., 2014). Examining these kinds of developmental outcomes around 18 months of age is crucial, given the rapid growth in vocabulary skills that happens around this time (Kuhl, 2010).

Despite much theory suggesting various pathways through which prenatal anxiety could impact child language development, research on the associations between these constructs is limited. Child language development is important to understand, given the crucial insight it may provide into other aspects of child development, such as cognitive and social abilities (Anderson & Freebody, 1981; Gertner et al., 1994).

Objectives and Rationale

The present research aims to examine the association between prenatal anxiety and child language development at 18 months of age. Relatively few published studies have examined the association between prenatal maternal mental health and language development, let alone the specific association between prenatal anxiety and language at 18 months. An important consideration for correlational research on this topic is the susceptibility to omitted variables bias (Duncan et al., 2004). Given this possibility, examining the links between prenatal maternal anxiety and child development with a full set of controls for potential confounders may yield a better sense of whether these associations reflect bias or a true association. It is important to understand that although various mental health disorders may be comorbid with one another, they present with different symptoms, require different therapeutic interventions, and–in the case of parent-child dyads–signify different implications for the child (Rogers et al., 2020).

Insofar as this correlational work could lay the foundation for testing whether interventions that target maternal mental health have positive effects on child language development, establishing whether these associations are robust is an important first step. By focusing specifically on prenatal anxiety, we can hone in on this construct as it relates to child language development. Furthermore, by controlling for numerous potentially confounding variables, including prenatal maternal depression, this research may strengthen the current research on this topic. Not only does the present research hope to lay the framework for possible interventions for mothers struggling with mental health, anxiety in particular, but it also strives to do so for children who may be susceptible to some of the unintentional consequences prenatal anxiety may bring. By distinguishing the potential implications of prenatal anxiety from those of prenatal depression and narrowing the scope, we enhance our understanding and ability to develop tailored interventions aimed at the prevention of any adverse developmental outcomes that may surface in the presence of prenatal maternal anxiety.

Hypothesis

The present research hypothesizes that higher levels of prenatal maternal anxiety are associated with lower levels of expressive and receptive language skills in children as measured at 18 months.

Methods

Participants

The mothers included in the current research were originally recruited for participation in a longitudinal study known as BUDDY, a study examining associations among various sociocultural factors, early experiences, and child development, particularly language, memory, and cognition. This study was approved by the Teachers College Institutional Review Board (IRB), and all participants provided written informed consent after being briefed on the study's purpose and procedures. They were also informed that they could refrain from specific measure collection or withdraw at any point. Participation was voluntary, though participants were compensated with differing cash amounts based on the time point. Participants were recruited in New York City through social media, community events, and prenatal clinics. An intentionally racially and socioeconomically diverse sample was recruited for participation. Recruitment for this study occurred in two waves, with a pause due to the COVID-19 pandemic. Thus, 93 mothers were recruited in the first cohort (before the pandemic) and 116 were recruited in the second cohort (after the height of the pandemic). In this study, we used data from both cohorts. Before enrollment in the study, mothers were screened over the phone to ensure that they met inclusion requirements, including that they were 18 years or older, were at least 35 weeks along in their pregnancy, and had no knowledge of fetal neurological or developmental issues. Upon successful enrollment in the study, all 209 mothers completed the prenatal visit. After the child was born, further inclusion criteria were enforced for subsequent participation. These criteria included the child's gestational age being greater than or equal to 37 weeks with no known neurological or developmental issues at birth. 205 mother-child dyads met these criteria. Participants were invited to participate in visits occurring from approximately every six months after birth through their child's third birthday.

In the current research project, we examined data from families who completed both the measure of prenatal anxiety and the measure of child language development at 18 months. The total number of mother-child dyads with complete data was 65 for cohort one and 104 for cohort two. Two mothers failed to complete demographic measures used as covariates in the analyses. Thus, my final analytic sample comprised 167 participants. Table 1 presents descriptive statistics corresponding to demographics and covariate variables.

Prenatal Anxiety

Prenatal anxiety was measured by the Beck Anxiety Inventory (BAI). The BAI is a 21-item questionnaire that helps in rating anxiety levels. It is a self-report assessment developed in 1988 that asks about common anxiety symptoms and their impact over a week (Rung, 2024). The questionnaire assesses short-term symptoms, primarily those that are physical. Sample items include: "Numbness or tingling" and "Unable to relax". For each question, there is a uniform set of four answer choices with scores ranging from zero to three. As a result, overall scores ranged from 0 to 63 characterized by grouping as follows: 0 to 7 = minimal anxiety; 8 to 15 = mild anxiety; 16 to 25 = moderate anxiety; and 26 to 63 = severe anxiety (Rung, 2024). While the BAI is not a diagnostic tool, this score can help mental health professionals determine the severity of anxiety. The measure shows high discrimination (between anxious and non-anxious persons) in clinical populations and high internal consistency (α = .93) and test-retest reliability as measured over a week (r_{ICC} = .84; Oh et al., 2018). However, it is important to note that the measure shows limited discriminant validity against depressive disorders (Oh et al., 2018).

Language Development

Child language development at 18 months was measured by the Language Environment Analysis (LENA) Snapshot. The LENA snapshot is a 52-question, parent-completed questionnaire that measures a child's expressive and receptive language skills (LENA Foundation, 2022). Sample items from the LENA snapshot include "Does your child spontaneously produce sentences that are 10 or more words in length?" and "Does your child understand concepts like 'least', 'most', and 'first'?" The questionnaire is norm-referenced to a sample of typically developing children who participated in a natural language study by its creators (LENA Foundation, 2022). The LENA system uses an algorithm to automatically score responses and compares the score to the normative dataset of the language development of other monolingual, North American English-speaking children at that age. Following this, the system generates a standard score that estimates the child's developmental language age, placing them at below or above their true age (LENA Foundation, 2022). The assessment has strong test-retest reliability, acquiring a correlational value of ~.93 over two years (LENA Foundation, 2022). This tool has also been correlated to various standardized language and cognitive assessments, including the Cognitive Adaptive Test and the Preschool Language Scale 4th Edition, both the expressive and receptive measures. These correlations attained a rounded value of 0.93 with significance (LENA Foundation, 2022). The LENA snapshot has proven to be efficacious as it relates to monitoring the progress of language development as well as helping to identify the probability of language delay (LENA Foundation, 2022). Especially given the multilingual nature of the participants recruited for BUDDY, it is crucial to mention that the LENA snapshot has demonstrated limited validity in multilingual contexts.

Covariates

Prenatal Depression

Prenatal depression was measured by the Patient Health Questionnaire Depression Scale (PHQ-8). The PHQ-8 is an 8-item scale used as a diagnostic and severity measure for depression (Kroenke et al., 2009). It is a self-report questionnaire that implements a Likert scale with responses for items ranging from zero ("not at all") to three ("nearly every day"). The questionnaire asks respondents to consider the presence of the items' corresponding symptoms during the previous two weeks (Arias-de la Torre et al., 2023). The items correspond to symptoms outlined in the DSM-IV as diagnostic criteria for major depressive disorder. Sample items for this measure include "I am sad all the time and I can't snap out of it" and "I am so sad or unhappy that I can't stand it". The PHQ-8 is scored by summing the respondents' scores to each of the items, with total scores closer to zero representing lower levels of depression and total scores closer to 24 representing higher levels (Arias-de la Torre et al., 2023). The tool displays high internal consistency reliability (α = .82) and construct validity (Pressler et al., 2011). For the purposes of BUDDY, the PHQ-8 was administered as opposed to the PHQ-9, which includes a ninth question on suicidal ideation, due to the study's desire to focus on the core aspects of depression while avoiding potential issues that may arise with Item Nine in non-clinical settings.

Demographic Variables

Demographic variables included child sex, child age at LENA snapshot completion, mother's race/ethnicity (with White mothers as the reference group), years of formal education completed by the mother (i.e., maternal education), and family income.

Data Analysis Plan

To examine the associations between maternal prenatal anxiety and child language development, we ran a series of three regression models using SPSS v29 (IBM Corp). First, we ran a simple linear regression model to test the relationship between prenatal anxiety and language development. Next, we conducted a multiple linear regression model to account for demographic covariates: child sex, child age at LENA snapshot completion, mother's race/ethnicity, maternal educational attainment, and family income. Finally, we ran a multiple linear regression model in which we included prenatal depression in addition to the demographic covariates introduced in model two. We tested this final model to probe the unique association between prenatal anxiety and child language development while holding constant the influence of prenatal depression, given the robust high correlation between depression and anxiety.

Results

Descriptive Statistics

Beginning with some discussion on descriptive statistics of the sample, the sample for the present study was relatively diverse. 44% of mothers in the sample identified as White and 55% identified as non-White, while 40% identified as Hispanic or Latino and 59% as not. As for family income and maternal education, the range for these variables was vast, though both distributions did skew towards the right (i.e., the higher end of the spectrum). On average, child age at LENA snapshot completion was 18 months, and the total score on the PHQ-8 had an average of five (M = 5) with a left-leaning distribution (i.e., the lower end of the spectrum) representing lower levels of prenatal depression for the mothers in the sample, on average. Regarding the BAI and prenatal anxiety, the average score was approximately nine (M = 9) with the data skewing left as well, representing lower levels of prenatal anxiety for the mothers included in the present study. Lastly, the total LENA snapshot score (i.e., child language development) skewed right toward higher scores and greater language abilities with an average score of 104 (M = 104).

Table 3 shows the correlations among variables included in my analyses. Of note, a low negative correlation was observed between prenatal anxiety and language development, and a moderate negative correlation was observed between prenatal depression and language.

Interestingly, in the context of the effect it had on results, a high positive correlation was observed between prenatal anxiety and depression, which likely impacted the results attained by the third regression, which included all the covariates.

Analyses

First, we ran a linear regression model to examine the relation between prenatal anxiety and language development at 18 months without the inclusion of any covariates. The model results are presented in Table 4, alongside all of the results produced by my analyses. The model's standard deviation indicated a small negative association (SD = -0.06), indicating that a one standard deviation increase in maternal anxiety was associated with a .06 reduction in language development. In raw units, this corresponded to a reduction of -0.15 units on the measure of language development. However, importantly, this association was imprecise and not statistically significant (p = .43), indicating that the association is not statistically distinct from a correlation of zero.

To examine the extent to which this small association was biased by child and family characteristics, we then ran a model that controlled for the following variables: child sex, child age at LENA snapshot, mother's race and ethnicity, maternal education, and income. The association between anxiety and language development became slightly larger but remained imprecise and statistically non-significant. Despite this slight increase, the first and second regression analyses produced consistently small estimates that hovered around 0, suggesting a near zero effect. Although the association remained statistically non-significant, interestingly, in this second regression analysis, the direction of the β value flipped from negative to positive (β = .03). On the other variables included in this regression model, notably, two attained statistically significant coefficients, child sex (β = .20, p = .01) and mother's race when specified to be Black (β = -.20, p = .03).

The last regression model we ran included the previously mentioned covariates in addition to prenatal depression. The inclusion of depression in the model substantially changed the association between prenatal anxiety and children's language development. Indeed, the association increased from near zero in earlier models (β = .20) and became nearly statistically significant by conventional standards (p = .06). This pattern of results looks a lot like what we might expect if suppressor effects were at play, leading me to be suspicious about the estimate. Such effects will be discussed further; however, they tend to occur with highly correlated variables when simultaneously input as predictors. Of note, there was an observed negative association between maternal depression and child language (β = -.25, p = .02) as well as an association between child sex and child language (β = .20, p = .007). And, as mentioned earlier, there was a strong positive correlation between prenatal anxiety and prenatal depression (r = .70).

Discussion

Overall, this study found little evidence to suggest an association between prenatal anxiety and child language development at 18 months of age.

In the first analysis, which most closely replicated existing literature due to its refrain from controlling for covariates, opposing results to current literature were observed. For example, Ibanez et al. (2015) explored the associations between prenatal depression and anxiety with children's cognitive outcomes without controlling for depression. In contrast to the current study, they found that prenatal anxiety was significantly associated with poorer cognitive development, as characterized by language, at three years old (Ibanez et al., 2015). Results also conflict with those of Rogers et al. (2020), which aimed to assess the association between prenatal depression and anxiety with various developmental skills, including language, during the first 18 years of life. Again, in this study, there was no control for the covariate prenatal depression, yet researchers found that prenatal anxiety was associated with diminished language abilities. This pattern is consistent with other prominent studies such as Irwin et al. (2020) and Clifford et al. (2022), which found higher prenatal anxiety to be associated with poorer receptive communication skills and postnatal maternal anxiety to be associated with child expressive language, respectively.

The antithetical nature of these results alone emphasizes the need to continue conducting research within this realm to better understand the implications of prenatal mental health on child language development and other facets of child development at large in order to design and implement tailored interventions geared towards positive outcomes for mothers and children as needed. One can hypothesize many reasons why this first analysis, and ultimately all of the analyses in this study, yielded null results. In the context of this particular study, one important consideration is the small sample size. The small sample size of 167 participants can pose a myriad of limitations for data analysis and interpretation. For starters, smaller sample sizes make it more difficult to ascertain the validity of a found effect, especially small effects. This is attributed to the lower statistical power often achieved by smaller sample sizes. Small sample sizes also yield unreliable results due to low generalizability, making it difficult to accurately attribute results to the broader population.

In the second regression analysis, we similarly found a statistically nonsignificant association between prenatal anxiety and child language development. However, as previously noted, a peculiar change occurred with the standardized coefficient value for this association from the first analysis to the second–the direction of this value shifted from negative to positive. Though we should be careful not to heavily interpret these coefficients, given that they were both very close to zero, this hypothetical interpretation opposes current literature in the field, highlighting the need for continued research. On top of this, it raises the question of whether there could be cases in which the experience of anxiety confers advantages to parent-child interactions that support child development. However, again, ultimately the estimate from this model was quite small and suggestive of no association. Interestingly, in this second regression, we did observe two statistically significant associations. An association was found between child sex and language development, a finding that aligns with a substantial amount of literature in the field, depicting significant gender differences in language acquisition spanning from early childhood through adolescence. Though empirical evidence is not robust and findings are mixed, numerous studies highlight a female advantage in various aspects of language acquisition in early childhood (Clarke-Stewart, 1973; Heister, 1982). We also observed a negative association between maternal race when identified as Black and child language development. This finding raises concern that warrants further examination, given its illustration of a vulnerable population that may benefit from interventions geared towards the promotion of early childhood language development. This association has been observed in other literature, such as Cho et al. (2007), which observed that at three years old, White children of White mothers had a lesser likelihood of presenting with language delays than non-White children and mothers (the majority of them Black). The consistent nature of this association highlights an exceedingly real equity issue that needs to be addressed to attain more positive outcomes for all children, regardless of their race or their mother's race.

Lastly, in the third linear regression, though it yielded a statistically insignificant relationship between total prenatal anxiety and language development, a notable change was observed in the estimate attained for the prenatal anxiety–child language development relationship. The p-value for this relationship changed significantly from the second analysis (p = .72) to the third (p = .06). This change, although there is a lack of association, represents a marked increase in the predictive power of prenatal anxiety. This compelling finding signifies what this research hypothesizes to be a suppressor effect induced by the control of prenatal depression. A suppressor effect in a regression occurs when controlling for a variable increases the predictive power of the model, as observed here. This represents the possibility of a rather complex relationship between prenatal anxiety and child language development that is somehow mediated by the presence of prenatal depression, should said relationship exist. Moreover, in this regression model, two statistically significant associations were observed: one between child sex and language development and another between prenatal depression and language development. The former of the two was also noted in the previous regression; however, the latter is specific to this regression, given that prenatal maternal depression was introduced here. A negative association between prenatal depression and child language development at 18 months was observed, representing a relationship characterized by a diminishing of language abilities in the presence of higher prenatal maternal depression. This coincides with existing literature which has found that at 12 months, children of mothers who were diagnosed with major depressive disorder while pregnant showcased lower language abilities compared to the control group (O'Leary et al., 2019).

Limitations

Myriad factors limit the conclusions of the current study. As previously mentioned, the small sample size is a large limitation of the current study for its potential to lead to difficulties in detecting small effects and unreliable results. It is also worth noting the two-cohort study design necessitated by the pandemic. As previously mentioned, participants were recruited at two different time points due to the pandemic, resulting in two cohorts. Recruitment for cohort one began before the pandemic and was interrupted once restrictions were imposed. Under "normal" circumstances where participants are recruited at different time points, a study faces the risk of the "cohort effect", which is described as the variation in characteristics of a group of people over time based on a shared experience (Caruana et al., 2015). This effect is all the more applicable in this study where participants of cohort two were recruited after a shared experience as life-changing as the COVID-19 pandemic. The cohort effect implies that different experiences attributed to a given period can impact the perceptions and responses of a participant in a research study (i.e., economic circumstances; Caruana et al., 2015). Meaning that participants of cohort two may showcase attitudes and behaviors that are highly attributed to their shared exposure to and experience with the pandemic. This phenomenon can ultimately introduce bias into the research because it signals a possible distortion of observed trends attributed to differences in shared life experiences that may have independently implicated the results.

Additionally, the lack of consistency in the child's age at which mothers completed the LENA snapshot serves as a limitation. Considering the correlation between prenatal anxiety and age at snapshot completion is low (r = .1), this limitation would be more pertinent if there were some sort of correspondence between prenatal anxiety and child age at snapshot completion, such that mothers with higher levels of prenatal anxiety were systematically completing the LENA snapshot at a different time than those with lower levels of prenatal anxiety. However, it remains an important factor to consider. The target age for the current study was 18 months; however, mothers with children as young as 16 months and as old as 21 months filled out the snapshot for the 18-month time point, contributing to the data the current study used. Though the mean age at snapshot completion was approximately 18 months (M = 17.98), the standard deviation was 0.895, signaling that the data is rather spread out. This standard deviation– though not exorbitantly high–is notable given that, as mentioned previously children's vocabulary is characterized by rapid growth at 18 months of age, with studies showcasing that infants may add up to 10-20 new words to their vocabularies daily (Fenson et al., 1993; Reznick & Goldfield, 1992). It is important to note that in the current study we controlled for child age, which addresses this concern.

The data was also limited by the measures used to capture the independent and dependent variables–the Beck Anxiety Scale and the LENA snapshot. Both measures relied on parent reporting, which introduces the potential for social desirability bias, where individuals answer questions in a way that might make them seem more "acceptable" (Paulhus, 2017). One specific limitation of the LENA snapshot includes what may be perceived as a limited scope, given that the snapshot primarily focuses on expressive and receptive language, failing to capture other parts of language, such as metalinguistic awareness and higher-order language skills (LENA Foundation, 2022). Additionally, currently, there is limited validity for multilingualism with the snapshot questionnaire, given that responses are compared to norms established by monolingual, North American English-speaking children. This fails to account for the nuances which multilingualism adds to language development and acquisition. A large portion of the study's participants, mothers and children alike, fall into this demographic, raising a concern regarding the accuracy of language abilities as captured by LENA snapshot scores. A specific limitation of the Beck Anxiety Inventory (BAI) is its low discriminant validity in distinguishing between symptoms of anxiety and depression (Muntingh et al., 2011). This raises a concern for the present study's aim to isolate the effects of anxiety, as BAI scores may not accurately capture anxiety independent of depressive symptoms.

The study's correlational design also limits its implications for drawing causal inferences. While we had causal theories about the links between maternal anxiety and child development, it's impossible to discern whether a causal relationship exists due to the lack of experimental manipulation in correlational research. Indeed, a range of confounding variables–including those that we did and did not control for–could have biased any observed associations between the constructs in our study. Though no significant association was observed in the present study, this limitation impacts all studies conducted without random assignment. Thus, it is important to consider how future studies may account for the inability to interpret correlation as causation.

Lastly, the current study is limited in its failure to account for or consider pregnancy-specific anxiety (PSA). Though often conceptualized interchangeably with prenatal anxiety, PSA is a unique phenomenon (Langille et al., 2023). PSA is common, affecting 20-40% of women during pregnancy (Araji et al., 2020). It differs from trait anxiety in that PSA is an affective state in which a pregnant person feels worry, fear, or nervousness surrounding matters related to their pregnancy such as their prenatal health, the baby's health, labor and delivery, and parenting (Dunkel et al., 2022). It is important to consider these phenomena distinctly considering they have distinct clinical impacts (Langille et al., 2023).

Conclusion

Despite the limitations mentioned above, the current study serves as a significant step in understanding the role that maternal mental health may play in child development, especially as it pertains to prenatal mental health and child language outcomes. Moreover, it contributes to the literature surrounding the importance of distinguishing between the ramifications of varying clinical disorders, which are crucial given their differences in treatment. Though the current study did not provide evidence that prenatal maternal anxiety influences child language development at 18 months, this null hypothesis suggests a need to achieve greater clarity into its true impact, given that these findings oppose current literature. The implications of improperly understanding the effects of prenatal anxiety, or maternal mental health, on child development are such that we remain incapable of providing adequate resources to children who may be at a disadvantage or to mothers who may be unaware of possible consequences. It also provides thoughtful insight into other associations present within this realm such as maternal race and child gender's seeming to influence child development as well. By means of attempting to answer a question and shed light on prenatal mental health, the current study raises new questions regarding patterns and relationships that may exist in the sphere of the mother-child matrix. It's important to understand the implications that may arise for child development based on prenatal maternal mental health to know how to best support pregnant mothers and their children not only during pregnancy but at other stages as well—including pre-conception, the postnatal period, and various time points throughout the child's lifespan.

Future Recommendations

The current paper posits a few recommendations for future research within the prenatal mental health-child development domain. As previously mentioned, a limitation of the current research is the small sample size. Thus, it is recommended that future studies replicate this research with larger sample sizes to be able to detect small effects and draw reliable conclusions from the observed results more confidently. This characteristic would also improve the representativeness of the dataset, allowing for the well-informed curation and tailoring of interventions.

Another recommendation for studies examining the impacts of prenatal anxiety alone is to emphasize and find a way to distinguish prenatal anxiety from what is known as pregnancy-specific anxiety. The current study does not account for these types of mental health that are vastly different in factors such as longitudinal course and prediction of postnatal mood disturbance (Blackmore et al., 2016). Most importantly, these ideas are different in their foundations. Pregnancy-specific anxiety encompasses feelings of nervousness or worry regarding personal health and appearance while pregnant, the baby's health, experience with the birthing process, and the healthcare system, and so on (Chandra & Nanjundaswamy, 2020), whereas prenatal state anxiety represents more general feelings of anxiety that happen to occur during pregnancy. Due to these differences, future research examining anxiety during pregnancy should individualize these two forms and assess their implications for not only children but also the pregnant person separately. The discourse surrounding prenatal mental health is minimal, let alone that of pregnancy-specific anxiety. However, existing research has found that higher levels of maternal pregnancy-specific anxiety are associated with lower inhibitory control in girls aged 6–9 years old and lower visuospatial working memory in girls and boys of the same age (Buss et al., 2011). Thus, preliminary research has found that child outcomes associated with maternal pregnancy-specific anxiety are distinguishable from those of prenatal state anxiety, highlighting the pertinence of upholding this distinction in continued research.

Lastly, another recommendation for a study of this nature, within the same vein as the aforementioned recommendation, would be to control for more covariates, such as prenatal stress. It is important to note that the study whose data the current study used also measured maternal stress at the prenatal time point. To advance the mission of dissecting varying facets of mental health (i.e., depression vs. anxiety vs. stress, etc.) in order to further understand their differences and implications, it is important to account for as many variables as possible, including prenatal stress. Future studies could assess the association between prenatal mental health and child outcomes by comprehensively accounting for different components of the former. To date, in the domain of child language development, research has found that the level of prenatal stress can account for variance in 2-year-old toddlers' productive and receptive language abilities (LaPlante et al., 2004), showcasing the ability of prenatal stress to dictate child language development in some capacity. Differentiating aspects of maternal mental health and understanding their possible implications for not only children but also for mothers is incredibly pertinent. This is particularly evident given the intersection between research and understanding as well as intervention and policy. For the sake of best supporting these communities, it's essential to appreciate the intricacies of factors that may affect them.

Ultimately, the current study contributes to an under-researched area of child psychology and development. Although the findings do not necessarily demonstrate a cause for concern, it is essential to continue researching in this domain to identify opportunities for action.

References

Anderson, R. C., & Freebody, P. (1981). Vocabulary knowledge. *Comprehension and teaching: Research reviews*, 77-117.

Araji, S., Griffin, A., Dixon, L., Spencer, S. K., Peavie, C., & Wallace, K. (2020). An overview of maternal anxiety during pregnancy and the post-partum period. *Mental Health Journal, 4*(4), 47–56. https://doi.org/10.29245/2578-2959/2020/4.1221

Arias-de la Torre, J., Vilagut, G., Ronaldson, A., Valderas, J. M., Bakolis, I., Dregan, A., ... & Alonso, J. (2023). Reliability and cross-country equivalence of the 8-item version of the patient health questionnaire (PHQ-8) for the assessment of depression: Results from 27 countries in Europe. *The Lancet Regional Health – Europe, 31*, 100659. https://doi.org/10.1016/j.lanepe.2023.100659

Blackmore, E. R., Gustafsson, H., Gilchrist, M., Wyman, C., & O'Connor, T. G. (2016). Pregnancy-related anxiety: Evidence of distinct clinical significance from a prospective longitudinal study. *Journal of Affective Disorders, 197*, 251–258. https://doi.org/10.1016/j.jad.2016.03.008

Brouwers, E. P. M., Van Baar, A. L., & Pop, V. J. M. (2001). Maternal anxiety during pregnancy and subsequent infant development. *Infant Behavior and Development, 24*(1), 95–106. https://doi.org/10.1016/S0163-6383(01)00062-5

Buss, C., Davis, E. P., Hobel, C. J., & Sandman, C. A. (2011). Maternal pregnancy-specific anxiety is associated with child executive function at 6–9 years of age. *Stress, 14*(6), 665–676. https://doi.org/10.3109/10253890.2011.623250

Caruana, E. J., Roman, M., Hernández-Sánchez, J., & Solli, P. (2015). Longitudinal studies. *Journal of Thoracic Disease, 7*(11), E537–E540. https://doi.org/10.3978/j.issn.2072-1439.2015.10.63

Chandra, P. S., & Nanjundaswamy, M. H. (2020). Pregnancy-specific anxiety: An under-recognized problem. *World Psychiatry, 19*(3), 336–337. https://doi.org/10.1002/wps.20781

Cho, J., Holditch-Davis, D., & Belyea, M. (2007). Gender and racial differences in the looking and talking behaviors of mothers and their 3-year-old prematurely born children. *Journal of Pediatric Nursing, 22*(5), 356–367. https://doi.org/10.1016/j.pedn.2006.12.004

Clarke-Stewart, K. A. (1973). Interactions between mothers and their young children: Characteristics and consequences. *Monographs of the Society for Research in Child Development, 38*(6–7, Serial No. 153), 1–108. https://doi.org/10.2307/1165928

Clifford, B. N., Stockdale, L. A., Coyne, S. M., Rainey, V., & Benitez, V. L. (2022). Speaking of state of mind: Maternal mental health predicts children's home language environment and expressive language. *Journal of Child Language, 49*(3), 469–485. https://doi.org/10.1017/S0305000921000131

Conger, R. D., Conger, K. J., Elder, G. H., Jr., Lorenz, F. O., Simons, R. L., & Whitbeck, L. B. (1992). A family process model of economic hardship and adjustment of early adolescent boys. *Child Development, 63*(3), 526–541. https://doi.org/10.1111/j.1467-8624.1992.tb01644.x

Dailey, S., & Bergelson, E. (2022). Language input to infants of different socioeconomic statuses: A quantitative meta-analysis. *Developmental Science, 25*(3), e13192. https://doi.org/10.1111/desc.13192

De Asis-Cruz, J., Krishnamurthy, D., Zhao, L., Kapse, K., Vezina, G., Andescavage, N., Quistorff, J., Lopez, C., & Limperopoulos, C. (2020). Association of prenatal maternal anxiety with fetal regional brain connectivity. *JAMA Network Open, 3*(12), e2022349. https://doi.org/10.1001/jamanetworkopen.2020.22349

de Diego-Balaguer, R., Martinez-Alvarez, A., & Pons, F. (2016). Temporal attention as a scaffold for language development. *Frontiers in Psychology, 7*, 44. https://doi.org/10.3389/fpsyg.2016.00044

Emmen, R. A., Malda, M., Mesman, J., van IJzendoorn, M. H., Prevoo, M. J., & Yeniad, N. (2013). Socioeconomic status and parenting in ethnic minority families: Testing a minority family stress model. *Journal of Family Psychology, 27*(6), 896. https://doi.org/10.1037/a0034693

Field, T. (2017). Prenatal anxiety effects: A review. *Infant Behavior and Development, 49*, 120–128. https://doi.org/10.1016/j.infbeh.2017.08.008

Gertner, B. L., Rice, M. L., & Hadley, P. A. (1994). Influence of communicative competence on peer preferences in a preschool classroom. *Journal of Speech, Language, and Hearing Research, 37*(4), 913–923. https://doi.org/10.1044/jshr.3704.913

Gilkerson, J., & Richards, J. A. (2009). The power of talk: Impact of adult talk, conversational turns, and TV during the critical 0–4 years of child development. *LENA Research Foundation*.

Hart, B., Risley, T. R., & Kirby, J. R. (1997). Meaningful differences in the everyday experience of young American children. *Canadian Journal of Education, 22*(3), 323.

Heister, G. (1982). Sex differences in verbal fluency: A short note. *Current Psychology, 2*, 257–260. https://doi.org/10.1007/BF02684518

Hobel, C., & Culhane, J. (2003). Role of psychosocial and nutritional stress on poor pregnancy outcome. *The Journal of Nutrition, 133*(5, Suppl), 1709S–1717S. https://doi.org/10.1093/jn/133.5.1709S

Hurtado, N., Grüter, T., Marchman, V. A., & Fernald, A. (2014). Relative language exposure, processing efficiency and vocabulary in Spanish–English bilingual toddlers. *Bilingualism: Language and Cognition, 17*(1), 189–202. https://doi.org/10.1017/S136672891300014X

Ibanez, G., Bernard, J. Y., Rondet, C., Peyre, H., Forhan, A., Kaminski, M., Saurel-Cubizolles, M. J., & EDEN Mother-Child Cohort Study Group. (2015). Effects of antenatal maternal depression and anxiety on children's early cognitive development: A prospective cohort study. *PLOS ONE, 10*(8), e0135849. https://doi.org/10.1371/journal.pone.0135849

Iruka, I. U., LaForett, D. R., & Odom, E. C. (2012). Examining the validity of the family investment and stress models and relationship to children's school readiness across five cultural groups. *Journal of Family Psychology, 26*(3), 359. https://doi.org/10.1037/a0028290

Irwin, J. L., Davis, E. P., Hobel, C. J., Coussons-Read, M., & Schetter, C. D. (2020). Maternal prenatal anxiety trajectories and infant developmental outcomes in one-year-old offspring. *Infant Behavior and Development, 60*, 101468. https://doi.org/10.1016/j.infbeh.2020.101468

Kannass, K. N., & Oakes, L. M. (2008). The development of attention and its relations to language in infancy and toddlerhood. *Journal of Cognition and Development, 9*(2), 222–246. https://doi.org/10.1080/15248370802022696

Keim, S. A., Daniels, J. L., Dole, N., Herring, A. H., Siega-Riz, A. M., & Scheidt, P. C. (2011). A prospective study of maternal anxiety, perceived stress, and depressive symptoms in relation to infant cognitive development. *Early Human Development, 87*(5), 373–380. https://doi.org/10.1016/j.earlhumdev.2011.02.004

Kroenke, K., Strine, T. W., Spitzer, R. L., Williams, J. B., Berry, J. T., & Mokdad, A. H. (2009). The PHQ-8 as a measure of current depression in the general population. *Journal of Affective Disorders, 114*(1–3), 163–173. https://doi.org/10.1016/j.jad.2008.06.026

Kuhl, P. K. (2010). Brain mechanisms in early language acquisition. *Neuron, 67*(5), 713-727.

Langille, C., Andreou, P., & Piccinini-Vallis, H. (2023). The prevalence of pregnancy-specific perinatal anxiety in Nova Scotia. *Maternal and Child Health Journal, 27*(7), 1127–1132. https://doi.org/10.1007/s10995-023-03639-y

Laplante, D. P., Barr, R. G., Brunet, A., Du Fort, G. G., Meaney, M. L., Saucier, J. F., Zelazo, P. R., & King, S. (2004). Stress during pregnancy affects general intellectual and language functioning in human toddlers. *Pediatric Research, 56*(3), 400–410. https://doi.org/10.1203/01.PDR.0000136281.34035.44

LENA Foundation. (2022, November 30). Developmental snapshot. LENA. https://www.lena.org/developmental-snapshot/

Lewis, A. J., Austin, E., Knapp, R., Vaiano, T., & Galbally, M. (2015). Perinatal maternal mental health, fetal programming and child development. *Healthcare, 3*(4), 1212–1227. https://doi.org/10.3390/healthcare3041212

Masarik, A. S., & Conger, R. D. (2017). Stress and child development: A review of the family stress model. *Current Opinion in Psychology, 13*, 85–90. https://doi.org/10.1016/j.copsyc.2016.05.008

Muntingh, A. D., van der Feltz-Cornelis, C. M., van Marwijk, H. W., Spinhoven, P., Penninx, B. W., & van Balkom, A. J. (2011). Is the Beck Anxiety Inventory a good tool to assess the severity of anxiety? A primary care study in the Netherlands study of depression and anxiety (NESDA). *BMC Family Practice, 12*, 66. https://doi.org/10.1186/1471-2296-12-66

Newland, R. P., Crnic, K. A., Cox, M. J., & Mills-Koonce, W. R. (2013). The family model stress and maternal psychological symptoms: Mediated pathways from economic hardship to parenting. *Journal of Family Psychology, 27*(1), 96. https://doi.org/10.1037/a0031112

Oh, H., Park, K., Yoon, S., Kim, Y., Lee, S. H., Choi, Y. Y., & Choi, K. H. (2018). Clinical utility of Beck Anxiety Inventory in clinical and nonclinical Korean samples. *Frontiers in Psychiatry, 9*, 666. https://doi.org/10.3389/fpsyt.2018.00666

O'Leary, N., Jairaj, C., Molloy, E. J., McAuliffe, F. M., Nixon, E., & O'Keane, V. (2019). Antenatal depression and the impact on infant cognitive, language and motor development at six and twelve months postpartum. *Early Human Development, 134*, 41–46. https://doi.org/10.1016/j.earlhumdev.2019.05.021

Paulhus, D. L. (2017). Socially desirable responding on self-reports. In *Encyclopedia of Personality and Individual Differences*. Springer, 1–5.

Reznick, J. S., & Goldfield, B. A. (1992). Rapid change in lexical development in comprehension and production. *Developmental Psychology, 28*(3), 406–413. https://doi.org/10.1037/0012-1649.28.3.406

Rogers, A., Obst, S., Teague, S. J., Rossen, L., Spry, E. A., Macdonald, J. A., Olsson, C. A., Skouteris, H., Kothe, E., & Hutchinson, D. (2020). Association between maternal perinatal depression and anxiety and child and adolescent development: a meta-analysis. *JAMA Pediatrics, 174*(11), 1082–1092. https://doi.org/10.1001/jamapediatrics.2020.2554

Romeo, R. R., Leonard, J. A., Robinson, S. T., West, M. R., Mackey, A. P., Rowe, M. L., & Gabrieli, J. D. (2018). Beyond the 30-million-word gap: Children's conversational exposure is associated with language-related brain function. *Psychological Science, 29*, 700–710. https://doi.org/10.1177/0956797617742725

Rung, R. (2024, June). What is the Beck Anxiety Inventory (BAI)? *Healthline*. https://www.healthline.com/health/anxiety/beck-anxiety-inventory#basics

Straughen, J. K., Misra, D. P., Divine, G., Shah, R., Perez, G., VanHorn, S., Onbreyt, V., Dygulska, B., Schmitt, R., Lederman, S., Narula, P., & Salafia, C. M. (2017). The association between placental histopathology and autism spectrum disorder. *Placenta, 57*, 183–188. https://doi.org/10.1016/j.placenta.2017.07.006

U.S. Department of Health and Human Services. (2010, September). Speech and language developmental milestones. *National Institute on Deafness and Other Communication Disorders*. https://www.nidcd.nih.gov/health/speech-and-language

Villar, J., Gunier, R. B., Tshivuila-Matala, C. O. O., Rauch, S. A., Nosten, F., Ochieng, R., Restrepo-Méndez, M. C., McGready, R., Barros, F. C., Fernandes, M., Carrara, V. I., Victora, C. G., Munim, S., Craik, R., Barsosio, H. C., Carvalho, M., Berkley, J. A., Cheikh Ismail, L., Norris, S. A., . . . Kennedy, S. H. (2021). Fetal cranial growth trajectories are associated with growth and neurodevelopment at 2 years of age: INTERBIO-21st Fetal Study. *Nature Medicine, 27*(4), 647–652. https://doi.org/10.1038/s41591-021-01280-2

Warren, E. J., & Font, S. A. (2015). Housing insecurity, maternal stress, and child maltreatment: An application of the family stress model. *Social Service Review, 89*(1), 9–39. https://doi.org/10.1086/680043

Table 1

Descriptives (Demographic Variables)

	% (*n*) or Mean (*SD*)	Min	Max
Mother's age at prenatal visit	32% (5.68)	19	46
Child sex			
Male	46%		
Female	54%		
Mother race			
White	44%		
Black or African American	21%		
Asian	10%		
American Indian/ Alaska Native	1%		
Other	23%		
Missing	1%		
Mother ethnicity			
Hispanic or Latino	40%		
Not Hispanic or Latino	59%		
Missing	1%		
Family income	$184,827 (320,823)	1	$2,563,501
Maternal education	16 (3)	6	22
Child age at LENA snapshot completion	18 (1)	16	21
Patient Health Questionnaire total score	5 (4)	0	17
Total (*N*)	167		

Note. This table shows descriptive statistics for the demographic variables. The total participant count for the current study is 167 children/families, as stated in the last row in the table. Maternal Age is reported in years. Family Income is reported in USD. Maternal Education is reported in years. Child age at LENA (Language Environment Analysis) snapshot completion is reported in months. Family income was log-transformed.

Table 2

Descriptives (IV & DV)

	Mean (*SD*)	Min	Max
Beck Anxiety Inventory total score	9 (7)	0	31
LENA snapshot total score	104 (18)	64	136
Total (*N*)	167		

Note. LENA is representative of Language Environment Analysis.

Table 3

Correlation Matrix

	1	2	3	4	5	6	7	8	9	10	11	12
1. Prenatal anxiety	-	-0.02	-0.17	0.16	-0.09	0.00	-0.05	0.04	-0.08	-0.10	0.11	**0.70**
2. Language development	-0.02	-	0.19	-0.28	0.08	-0.18	-0.14	0.26	0.12	0.25	-0.08	-0.21
3. Mom_White	-0.17	0.19	-	-0.47	-0.30	-0.07	-0.40	0.46	0.19	0.11	-0.27	-0.12
4. Mom_Black	0.16	-0.28	-0.47	-	-0.18	-0.04	0.00	-0.22	-0.22	-0.06	0.11	0.24
5. Mom_Asian	-0.09	0.08	-0.30	-0.18	-	-0.03	-0.23	0.24	0.24	-0.04	-0.05	-0.14
6. Mom_American Indian	0	-0.18	-0.07	0.04	-0.03	-	0.10	0.10	0.10	-0.09	0.09	0.06
7. Mom_ethnicity	-0.05	-0.14	-0.40	0.00	-0.23	0.10	-	**-0.53**	-0.26	-0.02	-0.03	0.05
8. Mom_education	0.04	0.26	0.46	-0.22	0.26	-0.11	**-0.53**	-	0.37	0.14	-0.05	-0.12
9. Family income	-0.08	0.12	0.19	-0.22	0.24	0.10	-0.26	0.37	-	0.06	0.09	-0.15
10. Child sex	-0.10	0.25	0.11	-0.06	-0.04	-0.09	-0.02	0.14	0.06	-	-0.05	-0.10
11. Child age at snapshot	0.11	-0.08	-0.27	0.11	-0.05	0.09	-0.03	-0.05	0.09	-0.05	-	0.14
12. Prenatal depression	0.40	-0.21	-0.12	0.24	-0.14	0.06	0.05	-0.12	-0.15	-0.10	0.14	-

Note. Bolded correlations indicate statistically significant correlations ($p < .05$). Construct names rather than measure names are used in the matrix.

Table 4

Correlation Matrix

	Regression #1		Regression #2		Regression #3	
	B (SE)	*p*	*B* (SE)	*p*	*B* (SE)	*p*
Covariates	-0.06 (0.19)	0.43	0.03 (0.18)	0.72	0.20 (0.25)	0.06
Demographic	-		*		*	
Controls						
Prenatal Depression	-		-		*	
					-0.25 (0.52)	0.02

Note. Bolded correlations are significant of highly correlated variables. Only correlations relevant to the current study and the overarching literature are included in the matrix. Construct names rather than measure names are used in the table.
* Covariate Included
– Covariate Not Included

Graduate Student Journal of Psychology
Fall 2025 - Vol. 25

Exploring the Mediating Role of Mentalization in the Relationship Between Attachment Styles and Suicidal Ideation in a Non-Clinical Pakistani Sample

Hamna Ali Khan, Haya Naseem, Zara Sohail, & Arshi Shabnam
Institute of Professional Psychology, Bahria University, Karachi, Pakistan

The aim of the current study is to determine the relationship between attachment styles, mentalization, and suicidal ideation using a quantitative correlational survey design in a non-clinical Pakistani sample. A purposive-convenience sampling technique was employed to recruit N = 295 adults (males, n = 81; females, n = 214) from Pakistan, aged 18 to 55 years (M = 23.07, SD = 5.37), using a Google survey. It was hypothesized that there is a relationship between attachment styles, mentalization, and suicidal ideation. Moreover, it was also hypothesized that there is a relationship between anxious attachment style and self-mentalization. To test the hypotheses, data were collected through three questionnaires: the Mentalization Scale (MentS), the Revised Adult Attachment Scale (RAAS), and the Suicidal Ideation Attributes Scale (SIDAS). The study's results were analyzed using SPSS 29.0.1.0. The findings revealed that attachment styles have a relationship with mentalization, self-mentalization (MentS-S), and suicidal ideation, whereas no significant correlation was found between mentalization and suicidal ideation. Only self-mentalization was found to have a negative correlation with suicidal ideation (r = −0.14, p = .01). Anxious attachment style was found to have a negative correlation with self-mentalization (r = −0.42, p = .01) and a positive correlation with suicidal ideation (r = 0.26, p = .01). The current findings underscore the importance of integrating emotional regulation strategies in therapeutic work with individuals with anxious attachment styles to prevent the development of suicidal ideation.

Keywords: attachment styles, mentalization, suicidal ideation

Suicidal ideation and attempts are growing concerns across Pakistani society, as literature highlights that 40% of Pakistani university students have reported a history of suicidal ideation, out of which 30% experienced suicidal ideation in 2018, explicitly indicating the need for the development of suitable interventions (Bibi et al., 2019). To comprehend the underlying mechanisms of suicidal ideation, it is necessary to investigate attachment style and mentalization as two significant predisposing psychological factors (Green et al., 2021). This study was conducted to explore the relationship between attachment styles, mentalization, and the increasing rate of suicidal ideation in a non-clinical Pakistani sample.

To contextualize the present study within broader psychological literature, the following sections aim to provide a brief overview of attachment styles, mentalization, and suicidal ideation. Each construct was analyzed in relation to its theoretical underpinnings, empirical evidence, and previously established associations with one another. This conceptual overview informs the rationale for the proposed mediation model that hypothesizes that mentalization may mediate the relationship between attachment styles an individual has and suicidal ideation in a non-clinical Pakistani population.

Attachment styles refer to a person's way of relating in intimate, receiving, and caregiving relationships, primarily with attachment figures such as parents, romantic partners, and children (Levy et al., 2011). Early caregiving relationships are fundamental to normative developmental processes in human beings (Hofer, 1994).

The evolutionary role that attachment plays extends far beyond providing physical protection to the child. A child's attachment responses are triggered when their environment makes them feel insecure in some way. It has been found that the aim of developing an attachment system is to create a feeling of safety and security. Thus, the attachment system is considered the primary regulator of emotional experience (Warren et al., 1997). Across an individual's lifespan, internal representations and attachment styles remain moderately stable, progressing into adulthood (Fraley, 2002).

Attachment styles have been conceptualized across a continuum between secure and insecure, drawing on Bowlby's (1969) foundational theory of secure attachment (Sheinbaum et al., 2015). A securely attached child typically has a sense of reassurance that their primary attachment figure will be consistently available to fulfil their needs (Main & Morgan, 1996). This secure attachment fosters a positive self-image and optimistic expectations of others, especially within close relationships, leading to higher relational satisfaction and overall well-being (Aronoff, 2012).

Ainsworth et al. (1971) further delineated insecure attachment into two primary categories: anxious and avoidant. These insecure attachment styles stem from inconsistent or unstable caregiving experiences during infancy and childhood and are often associated with difficulties in adult relationships and lower occupational functioning (Ainsworth et al., 1971). Insecure attachment is further categorized into insecure-avoidant and insecure-anxious subtypes.

Individuals who exhibit an insecure-avoidant attachment style tend to minimize their emotional reliance on their primary caregiver and avoid seeking proximity or comfort while distressed (Behren et al., 2007). As they grow older, these individuals tend to devalue their close relationships, become emotionally distant, and struggle with emotional regulation (Main & Morgan, 1996). Furthermore, they may express frustration and anger towards their caregivers as dysfunctional efforts to gain attention and connection (Solomon et al., 1995). In contrast, individuals with an insecure-anxious attachment style tend to be overly emotionally sensitive and have an anxious fixation with the availability of their primary caregiver. They tend to seek excessive reassurance and can become extremely distressed when their need for closeness is not consistently met (Cassidy & Berlin, 1994). The heightened activation of the attachment system often stems from inconsistency in caregiving during childhood, usually when the response to distress is unpredictable. As these individuals mature, they may struggle with feelings of abandonment, heightened need for reassurance, and emotional volatility in their personal relationships (Mikulincer & Shaver, 2007).

Mentalization, operationalized through the construct of reflective functioning, is defined as the ability to make sense of one's own and others' behaviors through the attribution of thoughts, feelings, and intentions (Fonagy et al., 2002). Mentalization has also been referred to as 'mind-mindedness', a process by which an individual understands that the mind mediates one's experience of the external world through the interpretation and the representation of different psychological states (Liotti & Gumley, 2009). Literature also defines mentalization as a deliberate stance that emphasizes interpersonal awareness, which brings about specific emotions, beliefs, intentions, and desires, subsequently producing a similar behavior (Liotti & Gumley, 2009).

Bateman and Fonagy (2006) identified three core dimensions of mentalization: the distinction between implicit and explicit functioning; the focus on two relational objects—the self and the other—and the integration of cognitive and affective aspects of both the content and process of mentalizing.

Implicit mentalization refers to the unconscious, involuntary, or procedural processes within an individual and their ability to imagine their own and others' various mental states (Bateman & Fonagy, 2006). For example, a parent responds to their infant's distress intuitively by soothing them without verbally reflecting on their child's state of distress (Fonagy et al., 2002). On the other hand, explicit mentalization is a process that is deliberately implied and used consciously. Explicit mentalization can be illustrated through the process of psychotherapy. The therapist and patient work collaboratively to consciously highlight and reflect upon the patient's mental states, encouraging the patient to deliberately imagine and focus on their cognitions, hence developing insight (Bateman & Fonagy, 2006). Self-mentalization emphasizes the importance of understanding and being aware of what an individual is feeling, as well as having insight into the way we react towards other individuals (Oldershaw et al., 2010). Self-mentalizing leads to the development of self-regulation, and thus, it plays a role in understanding our own individual feelings (Fonagy et al., 2002). Other-mentalization is defined as the ability to understand and interpret other individuals' mental states. An appropriate interpretation of others' mental states is linked to better social functioning (Perera & DiGiacomo, 2013) and a lower level of interpersonal conflicts (García-Sancho et al., 2014).

The third dimension of mentalization encompasses both cognitive and affective aspects, involving the content of mentalizing activity. The focus of this dimension is the 'intentional mental state' in the self and others, which is cognitively focused and affectively weighted to varying degrees. Mentalization requires the display of intact cognitive skills that enable an individual to imagine mental states with flexibility, plausibility, and complexity, while optimally integrating reason and insight with emotion. The integration of the affective and cognitive aspects of both the processes and content of understanding mental states helps individuals to "feel clear" and increases "emotional knowing" (Choi-Kain & Gunderson, 2008).

Literature indicates that the theory of mind, also referred to as mentalization, occurs within the prefrontal cortex (Fonagy et al., 2002). This brain region is associated with social intelligence and plays a vital role in controlling and processing behavior. The lateral fronto-temporoparietal network facilitates mentalization that is focused externally, enabling individuals to read facial expressions and posture. The bridge that connects affect regulation and social detection is primarily divided between the hypothalamus, amygdala, and bed nucleus of the stria terminalis, which gives emotional meaning to social cues and enables healthy individuals to react accordingly (Kateryna & Tanas, 2019).

To develop mentalizing skills, children must experience proximity and care from their primary caregivers early on (Meins et al., 2002). Research suggests that children's capacity to mentalize is closely intertwined with the quality of their attachment relationships (Fonagy et al., 2002). Primary caregivers aid them in understanding their own mental states and the mental states of others, which in turn helps the development of mentalization. However, if a child were to develop an insecure attachment, their capacity to engage in mentalization would be impaired (Fonagy & Luyten, 2009). Primary caregivers' constructive feedback helps the child reflect on their thoughts, feelings, and intentions (Sroufe, 2005). Self-reflection and reflective functioning are integral components of the interactions between children and their caregivers (Fonagy et al., 2002).

The development of mentalization is associated with emotion regulation, impacting certain aspects of an individual's functioning and possibly contributing to abnormal psychopathology. One product of this abnormal psychopathology can be the risk of suicidal ideation (Allen et al., 2003). Inability to understand mental states and insufficient regulation of emotions have been found to lead to problems in functioning (Allen et al., 2003).

A study by Venta and Sharp (2015) exploring the relationship between mentalization, attachment insecurity, and peer difficulties among adolescent inpatients found that mentalization mediated the relationship between disorganized attachment and peer problems. These findings highlight the central role of reflective functioning in shaping interpersonal outcomes and point to the potential relevance of mentalization as a psychological mechanism through which attachment-related vulnerabilities may influence broader emotional and relational difficulties (Venta & Sharp, 2015). In light of evidence linking peer problems and interpersonal dysfunction with a range of adverse mental health outcomes, these results offer a valuable framework for exploring how disruptions in mentalization may contribute to more severe internalizing experiences in vulnerable populations.

Mentalization is also influenced by the quality of specific attachment relationships (Bączkowski & Cierpiałkowska, 2015). Attachment avoidance within particular relationships has been found to be associated with impaired perspective-taking ability in those contexts, whereas general mentalizing capabilities were not significantly influenced by overall attachment quality. Therefore, the association between attachment and mentalization may be relationship-specific, with differences in reflective functioning being shaped by the emotional quality of each interpersonal bond (Bączkowski & Cierpiałkowska, 2015).

Suicidal ideation is an intricate public health issue that has its roots in social, etiological, biological, and psychological factors (Joe et al., 2008). Pakistan has a weak health system and even weaker mental health resources, as mental health is still widely considered a taboo due to the low literacy rate, financial disparity, and lack of awareness (Mashood, 2025). As a function of these deficits, attention is required to consider and address the alarming prevalence of suicidal ideation. Literature revealed that 35.6% of Pakistani students experienced suicidal ideation in the past, out of which 13.9% of students made a plan for death by suicide, and 4.8% of them died by suicide (Osama et al., 2014). Other similar studies show that 31.4% Pakistani students experience suicidal ideation, and there was no significant gender difference between males (29.2%) and females (33%) experiencing suicidal ideation (Khokher & Khan, 2005).

Adam (1994) explored the connection between attachment processes in suicidal behaviors, postulating that childhood attachment insecurities were considered a predisposing factor for developing suicidal ideation later in adolescence or adulthood. When an individual with an insecure attachment style faces loss or rejection, it leads them to an attachment crisis and distress, thus triggering the usage of maladaptive defense mechanisms. As a result, they become at high risk for indulging in self-destruction, self-harm, and developing suicidal ideation. In contrast, individuals who form a secure attachment with their caregivers tend to develop a more positive representation of not only themselves but also of others. When they experience interpersonal difficulties, they tend to be more resilient and manage distress better than those who have insecure attachment styles (Adam, 1994).

The Interpersonal Theory of Suicide explains that suicidal ideation is a result of low belongingness and perceived burdensomeness. It is this desire that leads towards the act of suicide (Orden et al., 2010). Rohani and Esmaeili (2020) aimed to elaborate on the susceptibility of suicidal ideation by examining the associations between psychological factors such as coping strategies, attachment styles, and dysfunctional attitudes. Their findings revealed that dysfunctional attitudes and insecure attachment styles function as stress-diathesis models in predicting increased susceptibility of suicidal ideation by affecting emotion-focused coping strategies (Rohani & Esmaeili, 2020). Empirical research has illustrated that attachment avoidance and anxiety are considered common risk factors for many psychological difficulties, including suicidal ideation (Sheftall et al., 2014). Furthermore, it has been found that those who have survived a suicide attempt reported significantly higher attachment avoidance and anxiety (Sheftall et al., 2014). When conditional logistic regression analysis controlled for family alliance and depression, suicide attempt status was predicted by attachment avoidance and anxiety (Sheftall et al., 2014).

Apart from deficits in attachment, deficits in mentalization are also independently linked to a higher risk of suicide. It is reported that compared to individuals with low risks of dying by suicide, individuals with moderate to severe risks of dying by suicide are 1.7 times more likely to report problems with mentalization (Pompili et al., 2020). Literature demonstrates that childhood trauma or distress can be directly associated with suicidal behaviors and indirectly associated through the pathways of attachment and mentalizing, indicating that an individual's insecure attachment style, along with an impaired mentalizing ability, can explain the link between childhood trauma and suicidal behaviors (Stagaki et al., 2022).

Theorized Pathways Between Mentalization, Attachment Styles, and Suicidal Ideation

Figure 1 illustrates the direction of the relationship between the study's variables. It is proposed that attachment styles can have an impact on one's ability to mentalize their risk of suicidal ideation. This can be attributed to early childhood attachment to the primary caregiver. If the attachment is not secure, when the child faces distress and feels insecure, it influences their mentalization. The primary caregiver facilitates the healthy development of mentalization (self, other, and motivation) by providing constructive feedback on the child's experience and motivating them to reflect upon and pay attention to their various feelings, thoughts, and intentions. This fosters more social and situational awareness and indicates a relationship between attachment and mentalization (Sroufe, 2005). Individuals who have developed an insecure attachment style with their primary caregiver tend to have a higher predisposition to various psychopathologies, such as suicidal ideation, which can further develop into suicidal behaviors (Adam, 1994).

Moreover, it was postulated that stress and threatening situations (proximal factors), as well as disruptions in the formation of secure attachment styles (distal factors) with the primary caregiver, disengage individuals' mentalization capabilities, which can elicit psychopathologies, such as suicidal ideation (Fonagy & Luyten, 2009). Hence, it creates a relationship between attachment styles, suicidal ideation, and mentalization.

The significance of determining the relationship between attachment styles, mentalization, and suicidal ideation is essential and multifactorial. Alarmingly, suicide is now a leading issue in Pakistan that requires immediate attention (Mashood, 2025). It is imperative to determine the underlying psychological mechanisms that serve as a precursor to suicidal ideation. There are many predisposing and distal factors identified that lead to the development of suicidal ideation among individuals; among them are attachment styles and mentalization. Hence, the researchers decided to test these variables with a non-clinical sample, trying to understand early indicators of suicide risk.

In times of stress, individuals with insecure attachment become predisposed to maladaptation of the mentalization capacity and, in turn, become vulnerable to suicidal ideation (Green et al., 2021). The primary purpose of this research is to shed light on the role of attachment styles in early childhood that is linked to the development of mentalization and the development of suicidal ideation among individuals in later life.

Furthermore, these variables have not been previously studied specifically in Pakistan. As suicidal ideation is on the rise, finding a relationship between attachment styles, suicidal ideation, and the relationship between mentalization might help to create insight, identify early precursors for suicidal intent, and develop strategies for the betterment of the community.

Considering the prevalence of suicidal ideation across all age groups in Pakistan and around the globe, this is a pivotal area of study (Mashood, 2025). Determining the relationship between these variables can help in having a better understanding and identifying early indicators of suicidal ideation and the predisposing mechanisms behind it.

Linking attachment styles with mentalization and suicide will help us understand their impact on the developmental period of the child. It will help us understand how early childhood trauma, neglect, and inconsistent care can impact their mentalization and eventually lead to psychopathology such as suicidal ideation. Moreover, the findings of this research can help in the development and implementation of effective therapeutic strategies that take into account attachment style and mentalization.

To accomplish the aforementioned research objectives, we hypothesized that:

1. There is a significant relationship between mentalization, attachment styles, and suicidal ideation.
2. Mentalization mediates the relationship between attachment styles and suicidal ideation.
3. There is a relationship between anxious attachment style and self-mentalization.

Method

A quantitative correlational survey design was carried out to determine the relationship between attachment styles, suicidal ideation, and mentalization in a non-clinical sample. The present study was conducted during the COVID-19 pandemic, a time when there were widespread restrictions on physical access to clinical settings and the vulnerable population. Moreover, due to the cultural stigma and taboo surrounding suicidal ideation in Pakistani society, a non-clinical sample, selected through purposive convenience sampling, was used in this study. This research included responses from 295 adults aged between 18 and 55 years old, approached through a Google survey. The researchers in this study were females, and data were collected from their social and academic groups, resulting in a disproportionate number of female respondents. Gender disproportion emerged as a consequence of the sampling method rather than intentional selection. Gender was not controlled for in this study. Individuals diagnosed with a mental health disorder or those who did not disclose a reason to consult a psychologist were not included in this study.

Measures

Mentalization Scale

The Mentalization Scale (MentS) is a 28-item self-report measure of self-related mentalization, motivation to mentalize, and other related mentalization, with higher scores suggesting a more sophisticated capacity for mentalization (Dimitrijević et al., 2017). The age range for this scale is 18 to 65 years old. MentS has three subscales which include self-related mentalization (MentS-S) (items: 8, 11, 14, 18, 19, 21, 22, 26) other-related mentalization (MentS-O) (items: 2, 3, 5, 6, 10, 12, 20, 23, 25, 28), and motivation to mentalize (MentS-M) (items: 1, 4, 7, 9, 13, 15, 16, 17, 24, 27). Mean and median scores were used to calculate the overall mentalization.

MentS-S measures one’s ability to mentalize or imagine mental states for one's own emotions and affect, MentS-O measures one’s ability to mentalize or imagine mental states of others’ emotions and affect, and MentS-M measures one’s internal motivation to mentalize or imagine mental states for one's own or others' emotions and affect. The MentS is suitable for measuring mentalization in both clinical and non-clinical samples. It utilizes a 5-point Likert scale that ranges from completely incorrect (1) to completely correct (5), with higher scores suggesting a more sophisticated capacity for mentalization. As determined by Nunnally (1967), the MentS has an acceptable reliability in clinical and non-clinical samples. The internal consistency of the scale, or Cronbach's alpha, is .84 for non-clinical samples and .75 for clinical samples (Dimitrijević et al., 2018).

Revised Adult Attachment Scale

The Revised Adult Attachment Scale (RAAS) measures interpersonal relationship attachments of individuals between the ages of 18 and 72 years old, using a 5-point Likert scale. The scale ranges from not at all characteristic of me (1) to very characteristic of me (5), with higher scores reflecting stronger tendencies within the respective attachment (DEPEND, CLOSE, or ANXIETY). Developed by Collins (1996), the RAAS is based on the earlier Adult Attachment Scale (Collins & Read, 1990). RAAS was normed using 406 students at the University of Southern California, and the age range for this test is 18–72 years. Revised Adult Attachment Scale is an 18-item scale that contains three dimensions: DEPEND, (items: 1, 6, 8, 12, 13, 17), CLOSE (items: 2, 5, 7, 14, 16, 18), and ANXIETY (items: 3, 4, 9, 10, 11, 15; Collins, 1996).

The DEPEND subscale measures an individual's belief that others can be available or relied upon when needed. The CLOSE subscale assesses an individual's comfort with intimacy and closeness. The ANXIETY subscale measures the extent to which a person feels anxious about elements such as abandonment or not being loved. The internal consistency of the RAAS subscales lies in an acceptable range (α = .78 for DEPEND, α = .77 for CLOSE, and α = .85 for ANXIETY; Collins, 1996; Nunnally, 1967).

Suicidal Ideation Attribute Scale

The Suicidal Ideation Attributes Scale (SIDAS) was developed using a sample of 1,352 Australian adults. The measure assesses the severity of suicidal ideation on a 10-point scale, with higher scores indicating greater persistence and severity of suicidal ideation, in individuals over the age of 18 years old (van Spijker et al., 2014). The reliability of the SIDAS lies within the acceptable range, exhibiting high internal consistency (α = .91) with good convergent validity (Nunnally, 1967). Suicidal ideation is classified as having ideas, plans, or thoughts of ending one's life. Individuals with scores that are greater than or equal to 21 are considered to have a high risk of suicidal behavior. The SIDAS consists of five items, each targeting an attribute of suicidal thoughts: frequency, closeness to attempt, controllability level of distress associated with the thoughts, and impact on daily functioning (van Spijker et al., 2014).

Procedure

From inception to the culmination of the data-collecting procedure, a rigorously functional hierarchy of supervision was maintained to ensure that all work involved in this research met high-quality standards, integrity, and the probability of errors was effectively minimized. Initially, a research proposal was designed, which explicitly outlined the research objectives to formally request authorization to collect data using the MentS, RAAS, and SIDAS.

After receiving approval from the university, a Google Survey was created based on the topic "Relationship between Attachment Styles, Mentalization, and Behavioral Tendencies" and disseminated across social and academic groups. Participants completed consent and demographics forms prior to responding to the three-scale questionnaire. The demographics collected included participants' gender, age, and socioeconomic class. All data were kept confidential.

Supervisors remained involved at every step of the data collection process. To incentivize participation, researchers included all participants' emails in a raffle. The emails of three participants were selected using a random picker tool to receive a gift basket or a one-month Netflix subscription. Upon selection, the winning participants were contacted via email and awarded one of two gift baskets, which included snacks or a one-month free Netflix subscription. The researchers made several ethical considerations, such as maintaining confidentiality by only discussing results with peers involved in the study. Furthermore, participants were informed of their right to withdraw from the study at any point. Additionally, no physical or mental harm was induced during the study. Although slight deception was used where "Suicidal Ideation" was masked as "Behavioral Tendencies" to minimize the chances of self-report bias, participants were debriefed following their participation.

Data Analysis

The data were first arranged in Microsoft Excel before being transferred to SPSS (Statistical Package for the Social Sciences) 29.0.1.0 for analysis. To explore the internal consistency, Cronbach's alpha coefficient was calculated for each scale used in the survey. Descriptive results include means, standard deviations, skewness, and kurtosis for the variables. Skewness and kurtosis were used to assess the normality of the data. Results indicated that all variables were normally distributed except for suicidal ideation. Then, inferential statistical tests, which included correlation analyses, were used to examine the strength and direction of relationships between key variables, with the significance level set to $p < .05$. To examine the simultaneous effect of mentalization, attachment styles, and suicidal ideation, a multiple regression analysis was conducted to find out their unique contribution to the variable of suicidal ideation. To investigate whether anxious attachment style is associated with self-mentalization, a simple linear regression analysis was conducted. This helped to determine the direction and evaluate the predictive relationship. Following the regression analysis, a mediation analysis was conducted where attachment style was entered as the independent variable, suicidal ideation as the dependent variable, and mentalization as a mediator. Bootstrapping was used to estimate direct and indirect effects with a confidence interval of 95%. All statistical tests conducted used a significance level of $p < .05$. The findings were then interpreted and discussed in detail, along with consideration of the study's limitations, recommendations for future research, and practical implications.

Results

The following tables represent the results of the study,illustrating the statistical relationship between the variables of attachment styles, mentalization, and suicidal ideation examined in a non-clinical Pakistani sample.

Table 1

Frequency and Percentages of Demographic Variables (N = 295)

Table 1 summarizes the demographic characteristics of the sample, including age, education level, and gender. The frequency and percentages of the demographic variables illustrate the composition of the sample.

Table 2

Descriptive Statistics and Alpha Reliability Coefficients, Univariate Normality of Study Variables (N = 295)

Table 2 illustrates the mean, standard deviation, skewness, and Cronbach's alpha reliability coefficient for each variable. The data demonstrate normal distribution for all variables except suicidal ideation. The reliability of all scales and subscales falls within an acceptable range.

Table 3

Correlations Between Attachment Styles (CLOSE, DEPEND, and ANXIETY), Mentalization (MentS-S, MentS-O, MentS-M), and Suicidal Ideation in the Sample of Pakistani Adults (N = 295)

Table 3 illustrates a positive weak correlation (r = 0.19) between attachment style and suicidal ideation. The table indicates a significant positive weak correlation between a close attachment style and mentalization (r = 0.18) and between a close attachment style and self-mentalization (r = 0.25). A close attachment style was additionally found to demonstrate a negative weak correlation with suicidal ideation (r = −0.13). Moreover, a dependent attachment style was found to have a significantly positive, extremely weak correlation with mentalization (r = 0.06) and a significantly positive moderate correlation with self-mentalization (r = 0.31). A significant negative weak correlation exists between a dependent attachment style and suicidal ideation (r = −0.14). Close and dependent attachment styles were found not to correlate significantly with others' mentalization and motivation to mentalize (r = .01 and r = .07 for close; $r < .001$ and r = −.05 for dependent). Furthermore, an anxious attachment style was found to have a significant moderate negative correlation with self-mentalization (r = −0.42) and a positive weak correlation with others' mentalization (r = −0.11). An anxious attachment style was also measured to have a significant, weak positive correlation with the motivation to mentalize (r = 0.23). An anxious attachment style demonstrated a significant positive weak correlation with suicidal ideation (r = 0.26). No significant correlation of mentalization (r = −0.06), others' mentalization ($r < .001$), and motivation to mentalize (r = .02) was found with suicidal ideation. However, self-mentalization was found to have a significantly weak negative correlation with suicidal ideation (r = -0.14).

Table 4

Multiple Regression Analysis for Attachment Style, Mentalization, and Suicidal Ideation

Table 4 indicates the impact of attachment style on mentalization and suicidal ideation. The R^2 = .03 revealed that the predictors explained 3% variance in the outcome variable (F [3.33, 5.38] = 0.03, $p < .05$).

Table 5

Simple Linear Regression Between Anxiety Attachment and Self-Mentalization (n = 294)

Table 6

Simple Linear Regression Between Anxiety Attachment and Self-Mentalization (n = 294)

Table 7

Regression Coefficients of Self-Mentalization on Anxiety Attachment

Tables 5, 6, and 7 demonstrate that the overall regression model was statistically significant (F [1, 293] = 63.57, $p < .001$), indicating that anxious attachment accounted for 4.4% of the variance in self-mentalization (R^2 = .0442). The results revealed that anxious attachment was a significant negative predictor of self-mentalization ($\beta = -.422$, $B = -.42$, $p < .001$). Therefore, findings suggest that individuals with an anxious attachment style tend to report low self-mentalization.

Table 8

Mediation Model for the Effect of Attachment Style on Suicidal Ideation Through Mentalization

Table 8 indicates a mediation model that revealed a significant direct effect of attachment style on suicidal ideation ($B = .20$, $SE = .06$, $p < .001$), whereas the indirect effect of attachment style and suicidal ideation through mentalization was not significant ($B = .00$, $SE = .00$, 95% CI [−.00, .02]).

This study shows a significant positive weak correlation ($r = 0.19$) between suicidal thoughts and attachment style. The association between attachment and suicidal ideation was not mediated by mentalization (indirect effect: $B = .00$, $SE = .00$, 95% CI [−.00, .02]). With a prediction value of 4.42%, the current study shows an association between self-mentalization and anxiety attachment type.

Discussion

The aim of the current study was to explore the relationship between attachment styles, mentalization, and suicidal ideation in a non-clinical sample. It was hypothesized that there is a relationship between attachment styles, mentalization, and suicidal ideation. More specifically, we proposed that there is a significant relationship between an anxious attachment style and self-mentalization. Furthermore, it was also hypothesized that mentalization would mediate the relationship between attachment styles and suicidal ideation.

The first hypothesis proposed that there is a relationship between attachment styles, mentalization, and suicidal ideation. Results show several weak correlations between attachment styles, mentalization, and suicidal ideation. Results indicate there is a weak positive correlation between closeness ($r = 0.18$) and mentalization. Primarily, this finding may indicate that mentalization, particularly self-mentalization, is associated with certain aspects of an individual's attachment style; however, given that the correlation was weak, other factors may also be precursors to this correlation. Consistent with previous research (Sroufe, 2005), the present study found that individuals with a secure attachment style were also skilled at mentalization of self. Literature also showed that the primary caregiver of a child facilitates the development of overall mentalization and specifically, the self-mentalization by understanding the child's subjective experience and creating a link by giving feedback on the particular experience (Bowlby, 1978). This can influence the culmination of children paying attention to what they are experiencing, reflecting upon it, and having an understanding of their own emotional or mental state. Moreover, a caregiver's regulatory assistance can help develop secure abilities in a child while coping in times of stress, which may serve as a protective factor later on in life (Bowlby, 1978). The findings of this study align with previous theoretical findings, further supporting the notion that early attachment experiences may contribute to the development of mentalization capabilities.

Similarly, it has been found that the parents' mentalizing style may also contribute to the child's mentalization development (Fonagy & Campbell, 2019). If parents have strong self-mentalization, they are likely to provide a secure base for the child, which is likely to lead to a secure attachment; hence, the child is more likely to develop a healthy overall mentalization as well as a healthy capacity to self-mentalize (Fonagy & Campbell, 2019). A positive relationship between children's mentalization and mothers' reflective functioning was seen in previous research (Rosso et al., 2015). In line with prior studies, the present research also found a weak correlation between attachment styles and self-mentalization, reinforcing that attachment style may be crucial for the development of mentalization.

A close attachment style was found to have a weak negative correlation with suicidal ideation ($r = -0.13$), and a significant negative weak correlation exists between dependent attachment style and suicidal ideation ($r = -0.14$). Literature showcased that individuals with secure attachment tend to have a positive perspective of others and themselves. Such individuals also develop a positive inner sense of self-worth and reassurance with their close figures (Hietanen & Punamäki, 2009). Thus, there is a chance that those who have consistent positive beliefs may adjust better to situations as compared to insecurely attached individuals who are detached from their significant figures (Hietanen & Punamäki, 2009). Along with this, individuals who have a secure attachment may develop a positive inner working model about themselves (Bowlby, 1988), consistent with the findings of this study.

The findings of this study may also support the model proposed by Bowlby (1988), postulating that closely attached individuals are more susceptible to asking for help when in a crisis or threatening situation, instead of feeling helpless or worthless, which, after long exposures, could turn into suicidal ideation later turning into action. Additionally, it can be anticipated that, as securely attached individuals might have a positive thought pattern and schema about themselves, they are more likely to seek professional help in comparison to anxiously attached individuals (Hong & Park, 2012; Simpson & Rholes, 2017). Thus, supporting the idea that individuals with secure attachment style may have a lower tendency to resort to suicidal ideation.

Furthermore, various studies reported that there exists an inverse statistically significant correlation between suicidal ideation and close or dependent attachment style (Ozouni-Davaji et al., 2013), and participants who have a secure attachment style with their fathers are less involved in sibling bullying and depression or suicidal ideation when compared to those who had an anxious or insecure attachment with their fathers, hence consistent with Bowlby's model (Bar-Zomer & Klomek, 2018).

There is a weak positive correlation between suicidal ideation and anxious attachment style ($r = 0.26$), as indicated by the results. Adam (1994) proposed that when an individual with an anxious or avoidant attachment style faces distress or threatening situations, they are unable to draw resources from their close interpersonal relationships in comparison to an individual who has formed a close attachment with their primary caregiver. Additionally, individuals with anxious or avoidant styles have an increased sensitivity to interpersonal threats like disappointment, rejection, and loss; this leads to increased and frequent activation of their attachment system (Adam, 1994).

Furthermore, individuals with an anxious attachment style lack adaptive strategies, which is why they resort to suicidal ideation, thinking, and behavior to elicit support and care from others (Adam, 1994). The findings of this research may support the developmental model provided by Adam (1994). Similar results were also found, postulating that those with anxious and avoidant attachment styles may face distress; they may feel overwhelmed, worthless, and hopeless. Hence, this may later contribute to the development of suicidal ideation and behavior (Kerns & Stevens, 1996). Additionally, literature states that the intensity of suicidal ideations and related behavior increases as the crisis upsurges in their life for those who are anxiously attached, which may be an explanation of the results of this research (Sheftall et al., 2014).

Durkheim (1951), in his theory of suicide, proposed that individuals who opt for egoistic suicide are unable to find their place and face numerous shortcomings in adjusting to their given situations, as they feel abandoned by society itself. Then gradually, they may resort to suicidal ideation and view suicide as the only solution to free themselves from the loneliness they are feeling (Tremblay, 2005). Other literature also provided similar findings, highlighting that two psychological states that are likely to increase the occurrence of suicidal ideation and suicidal behavior are social alienation and a sense of low belongingness (Joiner, 2007). It can be presumed that if a person does not form a secure attachment, there is a chance that they may not have a secure sense of belongingness, and this may increase the probability of suicidal ideation (Joiner, 2007).

Results indicate that self-mentalization has a significantly weak negative relationship with suicidal ideation ($r = -0.14$). Literature proposes that individuals who lie in the higher range of emotional dysregulation tend to have a low level of mentalization. The two domains of emotional dysregulation that show close association with mentalization are the inability to develop an understanding of emotional responses and a deficiency of emotional clarity. This leads to various issues, including the inability of a person to make sense of experience, acting out, the use of splitting, etc. As a process to cope with the overbearing emotional content, these individuals tend to show suppression of emotional experience, distortion, denial, and a sudden upsurge in impulsivity (Marszał & Jańczak, 2018). Disapproval and rejection of emotional responses are associated with low levels of mentalization, which are characterized by poor representation of experience, limited capacity of abstract and symbolic relation, and a lack of comprehensive emotional content (Marszał & Jańczak, 2018).

Another significant finding of the study is that no relationship was found between mentalization and suicidal ideation. It can be inferred that this may be due to the cultural non-acceptance of suicidal ideation (Ahmed et al., 2016). Suicide is considered a controversial topic in Pakistani culture, due to which individuals find it difficult to display their intent or ideation of suicide (Khan, 2025). This could be one of the factors for why no significant relationship between suicidal ideation and mentalization appeared in the statistical analyses, as individuals might be hesitant in answering questions regarding the topic. However, since this study ensured that the anonymity of the participants was a priority, it is a more likely conclusion that there was no correlation between the two variables, as the sample chosen is from a non-clinical population who may be better adjusted.

An additional explanation of the sample being less prevalent in suicidal ideation is that the majority of the participants were in their adulthood, which is a more stable time comparatively; hence, it may lead to the development of stable ideologies and thought processes, given the prefrontal cortex has developed and stabilized (Cash & Bridge, 2009). The sample was selected from the non-clinical population as a deliberate step by researchers, as they were keen to study the results in a population unaffected by mental health difficulties to see a general, holistic picture of the relationship between these variables.

Another explanation that may be attributed to there being a weak correlation between suicidal ideation and all other variables and a lower level of suicidal ideation in the overall sample is that suicide is considered a contentious topic in Pakistan. It is said to be a punishable act in Islam and Pakistan (Mashood, 2025). Hence, a lower prevalence of the variable has been seen in the sample, as individuals may be hesitant to talk about it and reveal their ideation. Furthermore, Section 325, which is a criminal law in the Pakistan Penal Code, 1860, and is practiced throughout Pakistan, states that "Whoever attempts to commit suicide and does any act towards the commission of such offense, shall be punished with simple imprisonment for a term which may extend to one year, or with fine, or with both" (Macaulay, 1860).

Results also showed a lower prevalence of suicidal ideation, which may be attributed to the possibility that, in the pandemic, individuals may have developed a beneficial coping mechanism; therefore, even in times of uncertainty, loneliness, and isolation, individuals did not resort to suicidal ideation. This is supported by recent studies revealing that healthy coping mechanisms and strategies were used by Pakistanis during COVID-19 to maintain their well-being, resulting in individuals' contributions to social welfare, new hobbies, exercise, eating healthy, and getting good sleep (Khan et al., 2021).

We further hypothesized that there is a significant relationship between anxious attachment style and self-mentalization. The findings confirmed a weak negative correlation between anxious attachment style and self-mentalization ($r = -0.42$). The development of mentalization is said to be correlated with different types of attachment systems, which are primarily associated with interactions with caregivers. During development, with the help of a caregiver, children learn to recognize and understand emotions. They do this through mirroring emotions around them and receiving an adequate affect response. These contingent representations created by the child's state of mind are shaped by the extent of security in attachment relationships, which in turn facilitates the development of children's abilities to mentalize (Fonagy & Target, 1996); hence, it can be concluded that those with an anxious attachment style may have low mentalization of self (Fonagy & Target, 1996). One of the predominant concerns of individuals with an anxious attachment style is that they are very fearful of being abandoned by someone they love, so they are continuously searching for signs that it might happen (Campbell & Marshall, 2011). The mentalization model proposed that the understanding that the caregiver has of the child's experience and the reaction given to them provides a model for the child to understand the emotional states within the world. This modelling ultimately leads to children learning to reflect upon and understand their states of mind. This progression from initial assistance to current independent observation of the self and others is dependent on a healthy, consistent, and reliable emotional interaction between the child and the caregiver. These healthy encounters occur only when close attachment is present.

In contrast, when early caregivers are unable to reflect on the child's state of mind or are unable to provide the modelling required for them to learn to perceive different states of mind, children do not receive the correct instructions they require to develop this extremely essential capability. Therefore, anxious attachment may impact the development of the capacity to mentalize (Allen et al., 2012). Moreover, literature suggests that individuals who have an insecure attachment style face difficulties in regulating their own emotions, have unsatisfactory relationships with their peers, and experience anxiety symptoms (Bosquet & Egeland, 2006). Given the results of the current study, a similar correlation has been found between self-mentalization and anxious attachment, where a significant, weak negative correlation has been found, pertaining to all the aforementioned reasons (Dimitrijević et al., 2018).

The third hypothesis was that mentalization will mediate the relationship between attachment styles and suicidal ideation. However, no mediation effect was found between the variables. This may be attributed to multiple underlying factors. Firstly, this was a self-report survey; thus, there is a chance that the mentalization scale used in the study might not have been sensitive enough to capture its role as a mediator, as the scale is newly developed and lacks extensive research based on it. Interviews or observation-based assessments or experiments might provide a better chance to find out the role of mentalization as a mediator in this study (Stagaki et al., 2022). Secondly, the study was conducted on a normal and non-clinical population; the sample characteristics, size, range, and variability might have affected the mediation effect. Thirdly, there is a chance that other psychological or cognitive processes may mediate the relationship between attachment style and suicidal ideation, such as mental pain (Lutzman & Sommerfeld, 2024) or self-criticism (Falgares et al., 2017), and several other variables, which were not explored in the study, may have confounded the mediation effect.

Furthermore, several reasons can be identified for weak correlations between the tested variables. Firstly, various sample characteristics, such as size, diversity, and representativeness, may have influenced the study's results. The study was conducted on a non-clinical population; therefore, it might have shown low suicidal ideation, thus affecting the other variables. Moreover, the study design was a self-report measure; therefore, it didn't capture the depth of these constructs as interview-based assessments, experiments, or observational assessments could have. Furthermore, there is a chance that mood, trauma, stress, or recent experiences could have influenced both attachment patterns and mentalizing capacities; hence, weak or no relationships were observed (Bateman & Fonagy, 2006).

This study holds significant relevance within the Pakistani cultural context, as its findings may contribute to increasing awareness of attachment patterns, mentalization capacities, and suicidal ideation, and help with the identification of early precursors of suicidal risk within non-clinical populations. The findings of this study show that there is a significant weak correlation between attachment anxiety and suicidal ideation. These findings can have positive implications in the Pakistani population, as those who have attachment anxiety can benefit from learning more about emotional regulation and coping strategies to prevent themselves from developing suicidal ideation. They may also benefit from interventions such as Mentalization-Based Therapy (MBT) or Attachment-Based Family Therapy (ABFT) that strengthen emotional regulation and reflective functioning within such individuals.

Moreover, this also provides insight into individuals struggling with such attachment styles, offering an opportunity for professionals to tailor therapy accordingly. It can also suggest that parents should be brought into the conversation to prevent further attachment issues, eventually leading to a decrease in suicidal ideation. Furthermore, it highlighted that attachment styles and mentalization can pose as early indicators for suicidal risk, and preventative strategies can be developed for those who struggle with insecure attachment styles or low mentalization.

Additionally, this finding offers valuable insights for counselors, clinicians, and educators, enabling them to tailor interventions to address attachment vulnerabilities. Furthermore, integrating psychoeducation about attachment styles, mentalization, and their relationship with suicidal intent into school and university counselling programs can help young adults build and develop resilience against suicidal ideation.

Several limitations were identified throughout this study. This was a correlational survey, making it difficult to draw causal interpretations of the data, and as the sample was recruited through a purposive-convenience sampling, it was difficult to understand the underlying mechanisms in-depth. Therefore, as causality could not be determined, it was unclear whether attachment styles influence suicidal ideation or if other unmeasured variables were at play.

Moreover, the study was conducted in Pakistani society, where suicide is viewed as a taboo topic. Hence, the possibility of response bias, including social desirability effects, could have led to an underreporting of suicidal thoughts as all of the tests used in this study were objective tests. Therefore, it is recommended to use projective tests in the future to capture an accurate understanding of suicidal ideation in Pakistan.

Another limitation of our study was that due to the pandemic, we ran an online survey that included a greater number of questions, so there is a possibility that participants might have felt fatigued or bored during the completion of the form, which could have produced biased answers. Moreover, due to the pandemic, the available sample to recruit from was the community around us, which could have led to bias within the results or demand characteristics. As individuals were able to self-refer to the study and enter without consideration, this is why we have mostly female data and a mean age of 23 years. This limited the generalizability of the results to other genders and older adults, and other demographic groups, potentially reducing the applicability of the results across a wider population. Therefore, it is recommended for future studies to conduct in-person surveys to eliminate bias. It is also recommended to conduct the study on an equal number of males and females to increase generalizability.

The scale of mentalization was not an indigenous scale; therefore, it is also recommended that an indigenous survey be developed to understand the capability of mentalization in the Pakistani population in detail, so that it can facilitate the improvement of mentalization-based interventions within the native population. Moreover, to completely capture the essence of the mediation effect of mentalization, interview-based assessments or experiments should be used.

As the researchers only chose a non-clinical population, there was no or very low suicidal ideation found in the sample; hence, it is recommended to replicate the study with individuals who are diagnosed or struggle with different mental health difficulties, giving a different perspective to the study itself.

Researchers could also choose to work with different age cohorts and test various attachment styles as well as types of mentalization, such as looking into intergenerational gaps and transfer of attachment. Since this research was only a cross-sectional study, a longitudinal study could be conducted to gain better insight into the relationship between different variables.

Researchers found a gender disproportion within the sample, as due to the sampling style, mainly females opted to answer the questionnaire. Future researchers could control the study for gender and see whether there is a difference between the genders and how that plays a role in suicidal risk.

Conclusion

The purpose of the present study was to test the relationship between attachment styles, mentalization, and suicidal ideation. Furthermore, the researchers wanted to find out if mentalization mediates the relationship between attachment styles and suicidal ideation. In conclusion, it was found that there is a partial correlation between the variables, and mentalization does not mediate the relationship between attachment styles and suicidal ideation. Attachment styles were found to have a relationship with mentalization, self-mentalization, and suicidal ideation, whereas no significant correlation was found between mentalization and suicidal ideation overall. Only self-mentalization was found to have a significant negative weak correlation with suicidal ideation. Anxious attachment style was found to have a significant moderate negative correlation with self-mentalization and a significant positive weak correlation with suicidal ideation.

References

Adam, K. S. (1994). Suicidal behavior and attachment: A developmental model. In Z. J. E. Adam & K. S. Adam (Eds.), *Clinical and developmental perspectives* (pp. 275–298). Guilford Press.

Ahmed, Z., Bhati, I., & Kausar, R. (2016). Suicide as a social dilemma: Evidence from Pakistan. *Advances in Social Sciences Research Journal*. Retrieved from https://www.researchgate.net/publication/337973411_Suicide_as_a_social_dilemma_Evidence_from_Pakistan

Ainsworth, M., Bell, S., & Stayton, D. (1971). Individual differences in strange-situation behavior of one-year-olds. In H. R. Schaffer (Ed.), *The origins of human social relations* (pp. 17–57). Academic Press.

Allen, J., Bleiberg, E., & Haslam-Hopwood, G. (2003). Mentalizing as a compass for treatment. *Bulletin of the Menninger Clinic, 67*(1). Retrieved from https://www.researchgate.net/publication/265029919_Mentalizing_as_a_Compass_for_Treatment

Allen, J. G., Lemma, A., & Fonagy, P. (2012). Trauma. In P. Fonagy & G. O. Gabbard (Eds.), *Handbook of mentalizing in mental health practice* (pp. 419–444). American Psychiatric Publishing, Inc.

Aronoff, J. (2012). Parental nurturance in the standard cross-cultural sample: Theory, coding, and scores. *Cross-Cultural Research, 46*(4), 319–347.

Bączkowski, B. M., & Cierpiałkowska, L. (2015). Mentalization within close relationships: The role of specific attachment style. *Polish Psychological Bulletin, 46*(2), 285–299. https://doi.org/10.1515/ppb-2015-0035

Bar-Zomer, J., & Klomek, A. B. (2018). Attachment to parents as a moderator in the association between sibling bullying and depression or suicidal ideation among children and adolescents. *Frontier Psychiatry, 9*. https://pmc.ncbi.nlm.nih.gov/articles/PMC5857527/

Bateman, A., & Fonagy, P. (2006). Mentalizing and borderline personality disorder. In J. G. Allen & P. Fonagy (Eds.), *The handbook of mentalization-based treatment* (pp. 185–200). John Wiley & Sons, Inc. https://doi.org/10.1002/9780470712986.ch9

Behren, K., Hesse, E., & Main, M. (2007). Mothers' attachment status as determined by the Adult Attachment Interview predicts their 6-year-olds' reunion responses: A study conducted in Japan. *Developmental Psychology, 43*(6), 1553–1567. https://doi.org/10.1037/0012-1649.43.6.1553

Bibi, A., Blackwell, E., & Margraf, J. (2019). Mental health, suicidal ideation, and experience of bullying among university students in Pakistan. *Journal of Health Psychology, 25*(12), 1957–1968. https://doi.org/10.1177/1359105319864276

Bosquet, M., & Egeland, B. (2006). The development and maintenance of anxiety symptoms from infancy through adolescence in a longitudinal sample. *Development and Psychopathology, 18*(3), 517–545. https://doi.org/10.1017/s0954579406060275

Bowlby, J. (1978). Attachment theory and its therapeutic implications. *Adolescent Psychiatry, 6*, 5–33. Retrieved from https://psycnet.apa.org/record/1982-00026-001

Bowlby, J. (1988). *A secure base: Parent-child attachment and healthy human development*. Basic Books.

Campbell, L., & Marshall, T. (2011). Anxious attachment and relationship processes: An interactionist perspective. *Journal of Personality, 79*(4), 775–810. https://doi.org/10.1111/j.1467-6494.2011.00723.x

Cash, S. J., & Bridge, J. A. (2009). Epidemiology of youth suicide and suicidal behavior. *Current Opinion in Pediatrics, 21*(5), 613–619.

Cassidy, J., & Berlin, L. J. (1994). The insecure/ambivalent pattern of attachment: Theory and research. *Child Development, 65*(4), 971–981. https://doi.org/10.2307/1131298

Choi-Kain, W. L., & Gunderson, G. (2008). Mentalization: Ontogeny, assessment, and application in the treatment of borderline personality disorder. *The American Journal of Psychiatry, 165*(9), 1127–1135. https://doi.org/10.1176/appi.ajp.2008.07081360

Collins, N. L. (1996). Working models of attachment: Implications for explanation, emotion, and behavior. *Journal of Personality and Social Psychology, 71*(4), 810–832. https://doi.org/10.1037/0022-3514.71.4.810

Collins, N. L., & Read, S. J. (1990). Adult attachment, working models, and relationship quality in dating couples. *Journal of Personality and Social Psychology, 58*(4), 644–663.

Dimitrijević, A., Hanak, N., Altaras Dimitrijević, A., & Jolić Marjanović, Z. (2018). The Mentalization Scale (MentS): A self-report measure for the assessment of mentalizing capacity. *Journal of Personality Assessment, 100*(3), 268–280. https://doi.org/10.1080/00223891.2017.1310730

Durkheim, E. (1951). *Suicide: A study in sociology* (J. A. Spaulding & G. Simpson, Trans.). Free Press. (Original work published 1897)

Falgares, G., Marchetti, D., De Santis, S., Carrozzino, D., Kopala-Sibley, D. C., Fulcheri, M., & Verrocchio, M. C. (2017). Attachment styles and suicide-related behaviors in adolescence: The mediating role of self-criticism and dependency. *Frontiers in Psychiatry, 8*, Article 36. https://doi.org/10.3389/fpsyt.2017.00036

Fonagy, P., & Campbell, C. (2019). Supporting the social triad. A commentary on "Keeping culture in mind: A systematic review and initial conceptualization of mentalizing from a cross-cultural perspective". *Clinical Psychology: Science and Practice, 26*(4), Article e12305. https://doi.org/10.1111/cpsp.12305

Fonagy, P., Gergely, G., Jurist, E. L., & Target, M. (2002). *Affect regulation, mentalization, and the development of the self*. Other Press.

Fonagy, P., & Luyten, P. (2009). A developmental, mentalization-based approach to the understanding and treatment of borderline personality disorder. *Development and Psychopathology, 21*(4), 1355–1381. https://doi.org/10.1017/S0954579409990198

Fonagy, P., & Target, M. (1996). Playing with reality: I. Theory of mind and the normal development of psychic reality. *The International Journal of Psychoanalysis, 77*(2), 217–233. Retrieved from https://psycnet.apa.org/record/1996-00761-001

Fraley, R. C. (2002). Attachment stability from infancy to adulthood: Meta-analysis and dynamic modeling of developmental mechanisms. *Personality and Social Psychology Review, 6*(2), 123–151. https://psycnet.apa.org/record/2002-12990-003

García-Sancho, E., Salguero, J. M., & Fernández-Berrocal, P. (2014). Relationship between emotional intelligence and aggression: A systematic review. *Aggression and Violent Behavior, 19*(5), 584–591. https://doi.org/10.1016/j.avb.2014.07.007

Green, J., Berry, K., Danquah, A., & Pratt, D. (2021). Attachment security and suicide ideation and behavior: The mediating role of reflective functioning. *International Journal of Environmental Research and Public Health, 18*(6), Article 3090. https://doi.org/10.3390/ijerph18063090

Hietanen, O. M., & Punamäki, R.-L. (2009). Attachment and early working alliance in adult psychiatric inpatients. *Psychology and Psychotherapy: Theory, Research and Practice, 82*(1), 107–123. https://doi.org/10.1080/09638230600808053

Hofer, M. A. (1994). Hidden regulators in attachment, separation, and loss. *Monographs of the Society for Research in Child Development, 59*(2–3), 192–207. https://pubmed.ncbi.nlm.nih.gov/7984161/

Hong, Y., & Park, J. (2012). Impact of attachment, temperament and parenting on human development. *Korean Journal of Pediatrics, 55*(12), 449–454. https://doi.org/10.3345/kjp.2012.55.12.449

Joe, S., Stein, D. J., Seedat, S., Herman, A., & Williams, D. R. (2008). Prevalence and correlates of non-fatal suicidal behavior among South Africans. *The British Journal of Psychiatry, 192*(4), 310–311. https://doi.org/10.1192/bjp.bp.107.037697

Joiner, T. E. (2007). *Why people die by suicide*. Harvard University Press.

Kateryna, K., & Tanas, Ł. (2019). Associations of attachment avoidance, anxiety and mentalization. *Polish Psychological Bulletin, 50*(1), 1–10. https://doi.org/10.13140/RG.2.2.17225.90721

Kerns, K. A., & Stevens, A. C. (1996). Parent-child attachment in late adolescence: Links to social relations and personality. *Journal of Youth and Adolescence, 25*(5), 577–590. https://doi.org/10.1007/BF01537388

Khan, A., Lodhi, F., Rabbani, U., Ahmed, Z., Abrar, S., Arshad, S., Irum, S., & Khan, M. (2021). Impact of Coronavirus disease (COVID-19) pandemic on psychological well-being of the Pakistani general population. *Frontiers in Psychiatry, 11*. https://doi.org/10.3389/fpsyt.2020.564364

Khan, M. M. (2005). Suicide and attempted suicide in Pakistan. *The Journal of the Pakistan Medical Association, 55*(6), 269–270. Retrieved from https://www.researchgate.net/publication/12968025_Suicide_and_Attempted_Suicide_in_Pakistan

Khokher, S., & Khan, M. M. (2005). Suicidal ideation in Pakistani college students. *Crisis, 26*(3), 125–127.https://doi.org/10.1027/0227-5910.26.3.125

Levy, K. N., Ellison, W. D., Scott, L. N., & Bernecker, S. L. (2011). Attachment style. In J. C. Norcross (Ed.), *Psychotherapy relationships that work: Evidence-based responsiveness* (2nd ed., pp. 377–401). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199737208.003.0019

Lutzman, M., & Sommerfeld, E. (2024). Mental pain as a mediator in the association between avoidant attachment and suicidal ideation among older men. *Aging & Mental Health, 28*(9), 1278–1285. https://doi.org/10.1080/13607863.2024.2323957

Macaulay, L. (1860). *Pakistan Penal Code (Act XLV of 1860)*. Retrieved from https://www.pakistani.org/pakistan/legislation/1860/actXLVof1860.html

Main, M., & Morgan, H. (1996). Disorganization and disorientation in infant strange situation behavior: Phenotypic resemblance to dissociative states. In L. K. Michelson & W. Ray (Eds.), *Handbook of dissociation: Theoretical, empirical, and clinical perspectives* (pp. 107–138). Plenum Press. https://doi.org/10.1007/978-1-4899-0310-5_6

Marszał, M., & Jańczak, A. (2018). Emotion dysregulation, mentalization and romantic attachment in the nonclinical adolescent female sample. *Current Psychology, 37*(4), 1081–1089. https://doi.org/10.1007/s12144-017-9573-0

Mashood, A. (2025). Contemporary mental health issues in Pakistan: Focus on forensic psychiatry. *Journal of Pakistan Psychiatric Society, 22*(1), 943–945. https://doi.org/10.63050/jpps.22.1.943

Meins, E., Fernyhough, C., Wainwright, R., Das Gupta, M., Fradley, E., & Tuckey, M. (2002). Maternal mind-mindedness and attachment security as predictors of theory of mind understanding. *Child Development, 73*(6), 1715–1726. https://doi.org/10.1111/1467-8624.00501

Mikulincer, M., & Shaver, P. R. (2007). *Attachment in adulthood: Structure, dynamics, and change*. The Guilford Press.

Oldershaw, A., Hambrook, D., Tchanturia, K., Treasure, J., & Schmidt, U. (2010). Emotional theory of mind and emotional awareness in recovered anorexia nervosa patients. *Psychosomatic Medicine, 72*(1), 38–43. https://doi.org/10.1097/PSY.0b013e3181c6c7ca

Osama, M., Islam, M. Y., Hussain, S. A., Masroor, S. M. Z., Burney, M. U., Masood, M. A., Menezes, R. G., & Rehman, R. (2014). Suicidal ideation among medical students of Pakistan: A cross-sectional study. *Journal of Forensic and Legal Medicine, 27*, 65–68. https://doi.org/10.1016/j.jflm.2014.08.006

Ozouni-Davaji, R. B., Valizadeh, S., & Nikamal, M. (2013). The relationship between attachment styles and suicide ideation: The study of Turkmen students, Iran. *Procedia-Social and Behavioral Sciences, 84*, 1475–1479. https://doi.org/10.1016/j.sbspro.2010.07.259

Perera, H. N., & DiGiacomo, M. (2013). The relationship of trait emotional intelligence with academic performance: A meta-analytic review. *Learning and Individual Differences, 28*, 20–33. https://doi.org/10.1016/j.lindif.2013.08.002

Pompili, M., Erbuto, D., Innamorati, M., Luciano, M., Sampogna, G., Abbate-Daga, G., Barlati, S., Carmassi, C., Castellini, G., De Fazio, P., Di Lorenzo, G., Di Nicola, M., Ferrari, S., Gramaglia, C., Nanni, M. G., Pasquini, M., Pinna, F., Poloni, N., Serafini, G., & Fiorillo, A. (2022). The relationship between mental pain, suicide risk, and childhood traumatic experiences: Results from a multicenter study. *The Journal of Clinical Psychiatry, 83*(4), 21m14176. https://doi.org/10.4088/JCP.21m14176

Rohani, F., & Esmaeili, M. (2020). Psychological factors of vulnerability to suicide ideation: Attachment styles, coping strategies, and dysfunctional attitudes. *Journal of Education and Health Promotion, 9*, Article 50. https://doi.org/10.4103/jehp.jehp_260_19

Rosso, A. M., Viterbori, P., & Scopesi, A. M. (2015). Are maternal reflective functioning and attachment security associated with preadolescent mentalization? *Frontiers in Psychology, 6*, Article 1134. https://doi.org/10.3389/fpsyg.2015.01134

Sheftall, A. H., Schoppe-Sullivan, S. J., & Bridge, J. A. (2014). Insecure attachment and suicidal behavior in adolescents. *Crisis, 35*(6), 426–430. https://doi.org/10.1027/0227-5910/a000273

Sheinbaum, T., Kwapil, T. R., Ballespí, S., Mitjavila, M., Chun, C. A., Silvia, P. J., & Barrantes-Vidal, N. (2015). Attachment style predicts affect, cognitive appraisals, and social functioning in daily life. *Frontiers in Psychology, 6*, Article 296. https://doi.org/10.3389/fpsyg.2015.00296

Simpson, J. A., & Rholes, W. S. (2017). Adult attachment, stress, and romantic relationships. *Current Opinion in Psychology, 13*, 19–24. https://doi.org/10.1016/j.copsyc.2016.04.006

Solomon, J., George, C., & De Jong, A. (1995). Children classified as controlling at age six: Evidence of disorganized representational strategies and aggression at home and at school. *Development and Psychopathology, 7*(3), 447–463. https://doi.org/10.1017/S0954579400006623

Sroufe, L. A. (2005). Attachment and development: A prospective, longitudinal study from birth to adulthood. *Attachment & Human Development, 7*(4), 349–367. https://doi.org/10.1080/14616730500365928

Stagaki, M., Nolte, T., Feigenbaum, J., King-Casas, B., Lohrenz, T., Fonagy, P., & Montague, P. R. (2022). The mediating role of attachment and mentalising in the relationship between childhood maltreatment, self-harm and suicidality. *Child Abuse & Neglect, 128*, Article 105576. https://doi.org/10.1016/j.chiabu.2022.105576

Tremblay, J.-M. (2005, February 2). Émile Durkheim, « De la définition des phénomènes religieux » (1897-1898) [Text]. *Les Classiques des Sciences Sociales*. http://classiques.uqac.ca/classiques/Durkheim_emile/annee_sociologique/an_socio_2/phenomenes_religieux.html

Van Orden, K. A., Witte, T. K., Cukrowicz, K. C., Braithwaite, S. R., Selby, E. A., & Joiner Jr., T. E. (2010). The interpersonal theory of suicide. *Psychological Review, 117*(2), 575–600. https://doi.org/10.1037/a0018697

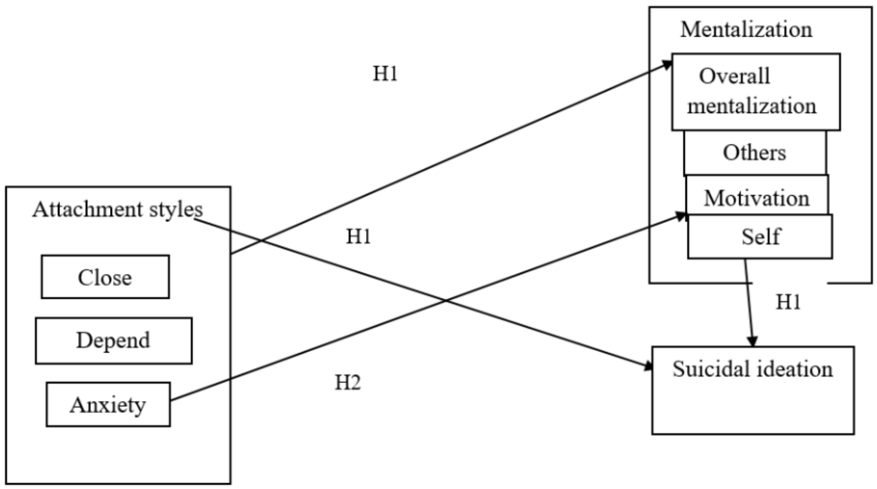
van Spijker, B. A. J., Batterham, P. J., Calear, A. L., Farrer, L. M., Christensen, H., Reynolds, J., & Kerkhof, A. J. F. M. (2014). The suicidal ideation attributes scale (SIDAS): Community-based validation study of a new scale for the measurement of suicidal ideation. *Suicide and Life-Threatening Behavior, 44*(4), 418–429. https://doi.org/10.1111/sltb.12084

Venta, A., & Sharp, C. (2015). Mentalizing mediates the relation between attachment and peer problems among inpatient adolescents. *Journal of Infant, Child, and Adolescent Psychotherapy, 14*(3), 323–340. https://doi.org/10.1080/15289168.2015.107199715.1071997

Warren, S., Huston, L., Egeland, B., & Sroufe, L. (1997). Child and adolescent anxiety disorders and early attachment. *Journal of the American Academy of Child and Adolescent Psychiatry, 36*(5). https://doi.org/10.1097/00004583-199705000-00014

Figure 1

Theorized pathways between the constructs of mentalization, attachment styles, and suicidal ideation



Note. The model draws links between the three hypotheses: (1) attachment styles, mentalization, and suicidal ideation are significantly interrelated; (2) mentalization functions as a mediating mechanism between attachment styles and suicidal ideation; and (3) anxious attachment is specifically associated with self-mentalization.

Table 1

Frequency and Percentages of Demographic Variables (N = 295)

Variables	*N*	%
Gender		
Male	81	27.5
Female	214	72.5
Socioeconomic Status		
Middle	101	34.2
Middle Upper	102	34.6
Upper Middle	82	27.8
Upper	10	3.4
Education		
Metric/O-levels	2	0.7
Intermediate/A-levels	67	22.7
Bachelors	171	58
Graduate	55	18.6
Consultation with a psychologist		
Yes	23	7.8
No	272	92.2
Diagnosed Mental Health Disorder		
Yes	0	0
No	295	100
Age		
18–55	295	100

Table 2

Descriptive Statistics and Alpha Reliability Coefficients, Univariate Normality of Study Variables (N = 295)

								Range	
Variables	*N*	Items	*α*	Mean	*SD*	*SK*	*K*	Actual	Potential
Suicidal	295	5	0.7	6.07	7.93	−0.08	0.19	0–43	0–50
Mentalization									
MentS-S	295	8	0.78	23.4	6.86	−0.04	−0.46	8–40	8–40
MentS-O	295	10	0.78	39.77	6.13	−0.87	1.52	15–50	10–50
MentS-M	295	10	0.68	38.65	6.02	−0.95	1.39	15–50	10–50
Attachment									
Styles									
CLOSE	295	6	0.63	3.15	0.75	−0.08	0.19	6–30	6–30
ANXIETY	295	6	0.85	3.35	1.03	−0.27	−0.76	6–28	6–30
DEPEND	295	6	0.53	2.55	0.67	−0.06	0.05	6–30	6–30

Note. MentS = Mentazlization, MentS-S = Mentalization of Self, MentS-M = Motivation of Mentalization, MentS-O = Mentalization of Others (**$p < 0.01$, *$p < 0.05$),, SK = Skewness, K = Kurtosis, Actual = Actual Range, Potential = Potential Range

Table 3

Correlations Between Attachment Styles (Close, Depend, and Anxiety), Mentalization (MentS_-S, MentS-O, MentS- M), and Suicidal Ideation in the Adult Sample of Pakistan (N = 295)

	CLOSE	ANXIETY	DEPEND	MentS	MentS-S	MentS-O	MentS-M	Suicidal
Attachment Style				−0.10				.19**
CLOSE	-	−.25**	.40**	.18**	.25**	.01	.07	−.13*
ANXIETY		-	−.46**	.14**	−.42**	.11*	.23**	.26**
DEPEND			-	−.06	.31**	< −.001	−.05	−.14*
MentS				-	.52**	.72**	.74**	−.06
MentS-S					-	−.04	< .001	−.14**
MentS-O						-	.54**	< .001
MentS-M							-	.02
Suicidal								-

Note. MentS = Mentalization, MentS-S = Mentalization of Self, MentS-O = Mentalization of Others, and MentS-M = Motivation of Mentalization

**$p < 0.01$, *$p < 0.05$.

Table 4

Multiple Regression Analysis for Attachment Style, Mentalization, and Suicidal Ideation

Variables	*B*	*SE*	*Df*	*F*	*T*	*p*
Constant	1.24	0.78			1.59	0.11
Attachment Style	0.20	0.62	1	3.33	3.23	< .001
Mentalization	−0.15	0.20	2	5.83	−0.75	.45

Note. ($p < 0.01$), *B* = Unstandardized beta coefficient, *SE* = Standard Error, *Df* = Degrees of Freedom, *F* = F-statistic, *t* = t-value, *p* = Significance Level

Table 5

Simple Linear Regression Between Anxiety Attachment and Self-Mentalization (N = 294)

Model	R	R^2	$Adj\ R^2$	Std. Error of the Estimate
1	.442[a]	.17	.17	6.23

Note. R = Correlation Coefficient, R^2 = Coefficient of Determination, $Adj\ R^2$ = Adjusted R-squared, [a] Predictors: (Constant), Anxiety Attachment

Table 6

Simple Linear Regression Between Anxiety Attachment, and Self-Mentalization (N = 294)

Model		Sum of Squares	*Df*	Mean Square	*F*	*p*
	Regression	2468.53	1	2468.53	63.57	.00[b]
1	Residual	11376.26	293	38.82		
	Total	13844.80	294			

Note. Sum of Squares = Total Variation, Df = Degrees of Freedom, Mean Square = Average Variation (Sum of Squares divided by Df), F = F-statistic, Sig. = Significance Level, [b] Predictors: (Constant), Anxiety Attachment

Table 7

Regression Coefficients of Self-Mentalization on Anxiety Attachment

	B	*SE*	*B*	*t*	*p*
Constant	32.79	1.23		26.59	< 0.001
ANXIETY	−.46	.05	−.42	−7.97	< 0.001

Note. B = Unstandardized Beta Coefficient, *SE* = Standard Error, *β* = Standardized Beta Coefficient, *t* = t-value, *p* = Significance Level

Table 8

Mediation Model for the Effect of Attachment Style on Suicidal Ideation Through Mentalization

Direct Effect						
					95% CI	
	B	*SE*	*T*	*P*	*LL*	*UL*
	0.20	0.06	3.23	< .001	0.07	0.32

Indirect Effect				
			95% CI	
	B	*SE*	*LL*	*UL*
	0.00	0.00	−0.00	0.02

Note. B = Unstandardized Beta Coefficient, SE = Standard Error, t = t-value, p = Significance Level, LL = Lower Limit of Confidence Interval, UL = Upper Limit of Confidence Interval

Caution when Crowdsourcing: Prolific as a Superior Platform Compared with MTurk

Daniel OConnell[1], Ashley Bautista[1], Clint Johnson[2], and Amanda Venta[1]
[1]*Department of Psychology, University of Houston, Houston, TX, USA*
[2]*Department of Psychology, Webster University, St. Louis, MO, USA*

Many researchers host surveys on online crowdsourcing platforms, such as Amazon's Mechanical Turk (MTurk) and Prolific. Online platforms promise a convenient way to meet sample size needs while drawing on diverse pools that might not otherwise participate in science. Yet, the quality of data obtained from these platforms is often questionable, so the collection must be closely monitored and reviewed. This study aimed to independently determine which crowdsourcing pool best serves researchers who plan to recruit for online surveys. To achieve this aim, we analyzed data from a recently completed study that drew participants from both MTurk and Prolific. We screened the collected data for both cost and quality, focusing on measures of attention, duration, and internal consistency. We found that only 9.89% of MTurk participants (N = 354) and 43.34% of Prolific participants (N = 345) produced high-quality data; Prolific also proved to be the more affordable option. Researchers considering these platforms for recruitment may weigh the evidence to make decisions when developing their own recruitment strategies. Finally, we highlight best practices for social scientists conducting online research, including additional survey and screening techniques.
Keywords: MTurk, Prolific, survey, crowdsourcing, data quality

Researchers have leveraged the internet for years, during which the use of crowdsourcing platforms has increased dramatically (Aguinis et al., 2021). Gosling and Mason (2015) extolled the use of the internet for research a decade ago, endorsing surveys conducted through crowdsourcing platforms to reduce costs and participant attrition. Moreover, it is just as easy now for researchers to collect survey data from undergraduates at their institution as it is to reach participants overseas, thereby reducing long-standing generalizability concerns (Best et al., 2001; Gosling & Mason, 2015). Online surveys bridge international borders; as of January 2023, 64.4% of the world's population was connected to the internet (Kemp, 2023). Yet, this approach to data collection has limitations, particularly regarding data quality. The aim of the present study was to investigate strengths of participant pools for social scientists to obtain high-quality data. To this end, we analyzed data from a completed study drawing participants from two major crowdsourcing platforms, MTurk and Prolific, and evaluated both data quality and cost.

History

Since 2005, Amazon's Mechanical Turk (MTurk) has promised to optimize efficiency, augment data collection, reduce researcher cost, and grant access to diverse participants (https://www.mturk.com/). Researchers (e.g., Aguinis et al., 2021; Smith et al., 2015) point to diverse participants, speed of data collection, and low cost as reasons for MTurk's widespread use. Yet, researchers have found that data quality and treatment of diverse populations on these platforms can suffer (e.g., Burnette et al., 2022). Indeed, Aguinis and colleagues (2021) highlighted ten areas in which MTurk is limited in its ability to collect high-quality data. These areas include participants lying about personal information (e.g., Webb & Tangney, 2022), lack of English fluency (e.g., Moss et al., 2021), and gathering data from professional survey takers (e.g., Cheung et al., 2017), all of which can reduce effect sizes (Chandler et al., 2015; Newman et al., 2021). The Webb and Tangney (2022) study serves as a provocative example of poor data quality collected via MTurk; just 14 of their 529 participants were reportedly "human beings" (p. 1). Webb and Tangney (2022) are not the only researchers to encounter of low-quality data from crowdsourcing platform participants (Bai, 2018; Simone, 2019; Stokel-Walker, 2018), and others have been critical of crowdsourcing from MTurk (Barends & Vries, 2019; Kennedy et al., 2020; Zack et al., 2019).

More recently, in 2014, Prolific came to the market, similarly guaranteeing a vetted, engaged, and more diverse participant pool from numerous countries with an emphasis on ethical pay (Peer et al., 2017; https://www.prolific.com/). Indeed, Prolific holds great potential to overtake MTurk as the optimal crowdsourcing platform (Palan & Schitter, 2018). Yet, direct comparisons between the two have yielded conflicting–and sometimes biased–results. For instance, Peer et al. (2017; 2022) portrayed Prolific as superior in terms of participant attentiveness, comprehension, honesty, and reliability compared to MTurk and CloudResearch, but both studies were funded by the Prolific company. Conversely, Litman et al. (2021) responded in a paper sponsored by CloudResearch, a company that accesses MTurk participants and aims to improve upon Amazon's platform. Their results demonstrated superior data quality on MTurk when paired with the CloudResearch Toolkit. Given directly contrasting results, it is important to establish objective criteria to fairly compare MTurk and Prolific recruitment, including data quality and pricing.

Deciding between MTurk and Prolific

Cost

One basic and practical consideration is cost. Prolific mandates that researchers pay their participants an ethical wage (Newman et al., 2021), which is a minimum of $8 U.S. Dollars (USD)/hr (https://www.prolific.com/). Meanwhile, MTurkers earn a minimum pay of $0.01 USD per assignment (https://www.mturk.com/). Prolific charges a higher platform usage fee (25% base rate for academics) compared to MTurk (20% base rate). However, Prolific includes most participant specifiers (e.g., age or job) within their base cost, whereas MTurk requires researchers to pay additional fees. Therefore, when participants are compensated equally, MTurk is cheaper until specifiers are added for researchers recruiting a specific population (e.g., young adults). Since MTurk does not enforce a minimum wage, researchers may pay participants less. Crump et al. (2013) found that higher wages did not incentivize participants enough to provide higher-quality results, but it did result in lower dropout rates. Conversely, Litman et al. (2015) showed that monetary compensation is a primary driver for participation, tying data quality to compensation rates, thus directly contradicting findings by Crump and colleagues (2013).

Data Quality

Researchers also value the quality of their data when using crowdsourcing platforms. Data quality is a term comprising many factors (Douglas et al., 2023), operationalized herein as–on the high-quality end–higher rates of passing attention checks and task completion combined with lower rates of lying and deception. Multiple methods are often combined to make conclusions about data quality (Douglas et al., 2023). Some techniques–often used in tandem with others–include attention checks, survey duration, and internal consistency.

Most studies employ attention checks (Douglas et al., 2023). To evaluate attention, survey designers may ask participants to make a forced response, write an open-ended response demonstrating understanding, or perform unrelated tasks like math–though they vary in effectiveness (Abbey & Meloy, 2017). That said, checks like these are not without detractors. Hauser et al. (2018) demonstrated that manipulation checks can confound results, particularly when implemented incorrectly (e.g., attention question placement is not randomized).

Another indicator researchers can use to determine data quality is survey duration (Teitcher et al., 2015). By comparing individual participant survey durations to the average and pilot data, researchers can identify outlier durations (Matjašič et al., 2018). Participants who respond far too quickly can be identified as suspicious and of low quality (e.g., Goodrich et al., 2023).

A third way to evaluate data quality is through internal consistency (e.g., Douglas et al., 2023). One way to evaluate internal consistency is through Cronbach's alpha (α) (Cortina, 1993), as random responding contributes to low values (Fong et al., 2010). When values are low (see Cortina, 1993), especially compared to validated standards of a measure, researchers should be skeptical about the overall reliability of their data.

Previous MTurk and Prolific Comparisons

A few independent studies have been conducted to directly compare MTurk and Prolific, demonstrating Prolific as superior. Albert and Smilek (2023) observed greater disengagement among MTurk participants compared to those on Prolific, though they only included high-performing MTurk users. While using participants identified by the platforms as high-quality can be beneficial for getting attentive participants (Lu et al., 2022), it limits random selection and naive respondents–those who are unfamiliar with certain measures (Matthijsse et al., 2015). In another direct comparison, Douglas and colleagues (2023) conducted an independent analysis across MTurk, Prolific, CloudResearch, SONA, and Qualtrics with a well-powered 500 participants per pool. They concluded that Prolific and CloudResearch outperformed the other pools in terms of data quality, with no substantial differences between the two; both outperformed the unmodified

MTurk. They also highlight other relevant details, such as the price per quality participant, wherein Prolific ($1.90) was cheaper than CloudResearch ($2.00) and MTurk ($4.36). Yet, similar to Albert and Smilek (2023), Douglas et al. limited participants by only allowing those who had already completed 100 surveys, thereby rejecting naive participants. The authors further suggest that their results ought to be regularly replicated, as pool demographic compositions fluctuate over time. The present study builds on these prior works by directly comparing MTurk and Prolific without pre-established participant quality standards.

Current Study

The current study aimed to directly compare the quality and cost of data gathered from identical surveys posted on MTurk and Prolific. Most previous studies comparing MTurk have pre-screened for high-performing users, limiting naive participants. In contrast, our study compared recruitment between MTurk and Prolific with naive and non-naive participants, representing the recruitment efforts commonly seen in contemporary research. Secondary data analyses were conducted on data collected in a previously completed study. Ultimately, we sought to answer the research question: How do cost and data quality from participants recruited from MTurk and Prolific differ without pre-screens in place? This question was answered using a thorough screening process influenced by prior research crowdsourcing data quality.

Method

Participants

For the MTurk sample (n = 354), most participants were White (81.64%), heterosexual (82.49%), and male (61.30%), with an average age of 26.18 years (SD = 4.54). For the Prolific sample (n = 345), most participants were White (77.08%), heterosexual (63.03%), and female (67.05%), with an average age of 22.20 years (SD = 2.03).

Procedures

This study utilized data collected through Qualtrics on MTurk and Prolific platforms. The current study aimed to compare samples drawn from MTurk and Prolific for a broader study (see more https://osf.io/2n8ge), which was approved by the IRB at Saint Louis University. Two identical surveys–differing only by the inclusion of an ID number for MTurk participants–were launched on the morning of April 15, 2022. Inclusion criteria required participants to be English-speaking young adults aged 18-25 and living in the United States.

Participants were told that they would be providing the company 'OCEAN' with feedback on their newly developed dating application rooted in personality. In reality, the study aimed to investigate participant preferences for romantic partners based on perceived personality and weight. Nevertheless, we subjected participants to a realistic process of testing a dating app which allowed them to create an OCEAN profile, rate eight random profiles, provide qualitative and quantitative feedback on the "app," and rate 34 images as high or low in BMI/weight. All participants were compensated $2 USD for approximately 15 minutes of work ($8 USD/hr rate).

Measures

Big Five Factor Model of Personality

The Mini-IPIP (Donnellan et al., 2006), a measure based on the Big Five Factor Model of Personality (Goldberg, 1999), was included as a component of the profile-building process to assess personality and induce psychological realism. The Mini-IPIP has demonstrated strong validity and internal consistency as a personality inventory (Donnellan et al., 2006). This measure was used to compare internal consistency before and after the screening process through Cronbach's α levels.

Demographics

Demographics were gathered through the profile-building process. Data included age, race, gender identity, sexual orientation, height, weight, and marital status.

Duration Data

Total survey duration captured via Qualtrics was used to compare quality before and after the screening process. Based on pre-launch trials, participants were expected to take a maximum of 15 minutes to complete the survey.

Data Quality Screening Process

The data screening process was inspired by the Webb and Tangney (2022) study, wherein participants were screened out in a step-by-step process and removed from the participant pool. The calculations for the cost of each high-quality respondent were inspired by Douglas et al. (2023).

The sequential screening process consisted of four steps: (1) age, (2) self-reported seriousness, (3) sensible open-ended responses, and (4) other sensible responses.

Participants outside the age inclusion criteria between 18-25, inclusive, were screened out. Then, the final question of the Qualtrics survey asked participants: "How seriously did you take this survey?" Responses ranged from 1-5, with 1 being "*not very serious*" and 5 being "*very serious*." Those who admitted to not taking the survey seriously were screened.

Two open-ended questions were analyzed to screen for unreasonable and duplicate responses. One of these questions asked participants to "Please briefly summarize the purpose of this survey," following the consent form (on a separate page). The second, towards the end, asked participants to "Provide any remaining thoughts on OCEAN here." Criteria for what was considered reasonable were developed *a priori* using manifest content analysis (Graneheim et al., 2017). Responses that were marked correct must have mentioned the words "develop," "personality," "test," "algorithm," or "dating app" and sufficiently explain the purpose of the study. Exactly identical response featured exactly the same words, spelling, capitalization, and punctuation were also screened out.

Two additional metrics were used to refine participant quality based on congruence. First, participants were asked to rate 34 images as high or low in BMI (https://osf.io/2n8ge). Two images (one male and one female) were presented twice to measure consistency. Second, participants who provided impossible heights and weights were screened.

Results

Data Quality on MTurk versus Prolific

Results from the screening process are summarized in Table 1 and explained below.

Age

Of the 354 MTurk and 345 Prolific participants, 125 of the MTurk participants reported an age outside the restricted age range on the survey. This left 229 (64.69%) MTurk and 345 (100%) Prolific participants for analysis, totaling 82.12% of the sample.

Seriousness

Two MTurk participants did not respond to this question, and one individual on the Prolific survey reported a rating of 2, meaning they did not take it seriously. This left the participant count at 227 (64.12%) for MTurk and 344 (99.71%) for Prolific, or 81.69% of the total.

Sensible Open-Ended Responses

About a third (113) of the remaining Prolific participants were removed for illogical or incorrect responses on one or both of the open-ended questions. An example of this type of response included, "the whole body of salt water that covers nearly three-fourths of the earth." As a result, 80 (22.60%) MTurk participants and 231 (66.96%) Prolific participants remained, or 44.35% of the total sample. Next, identical responses were removed. For example, the response "OCEAN developers to improve the algorithm of their new dating app." appeared three times on MTurk. This affected participants in both pools such that 64 (18.08%) MTurk and 229 (66.38%) Prolific participants, or 41.92% of the total, remained.

Other Sensible Responses

First, participants were screened for inconsistent responses to identical questions. Of the remaining participants, just 38 (10.73%) MTurk and 153 (44.35%) Prolific participants, or 27.32%, were consistent in rating both sets of images at this stage. Next, participants were screened for impossible heights and weights. This affected three participants on MTurk for entering: (1) 8 feet 8 inches while weighing 120 pounds, (2) a height of 1 foot 1 inch tall, and (3) a weight of 154324 pounds. After this step, 35 (9.89%) MTurk and 153 (44.35%) Prolific participants remained, representing 188 of the initial 699 (26.90%).

Internal Consistency

Table 2 compares internal consistency on the Mini-IPIP between MTurk and Prolific alongside the original psychometric study (Donnellan et al., 2006). Both the Prolific and MTurk α values improved substantially after screening. Although Prolific scores generally began higher, both the MTurk and Prolific pre-screen data would be considered unreliable (Cortina, 1993). Moreover, after screening, all of the α values were higher for Prolific except for Intellect/Imagination. As a result, the evidence would support the post-screen reliability in Prolific but not MTurk due to values below .70 (Cortina, 1993).

Duration

Total survey duration was used to compare quality before and after the participant screening process. Based on pre-launch trials, participants were expected to take up to 15 minutes to complete the survey. The times that participants took on MTurk before (*Mdn*

= 10 minutes and 20 seconds) and after (*Mdn* = 10 minutes and 18 seconds) screening were slightly longer than the times that participants took on Prolific before (*Mdn* = 8 minutes and 47 seconds) and after (*Mdn* = 9 minutes and 2 seconds) screening. Using 2 *SD*s from the mean in each sample as a metric to compare speed (Matjašič et al., 2018), no responses on either MTurk or Prolific were considered outliers in the "fast" direction. While a handful of slow outliers were present, this was not meaningful to this study, as participants had the freedom to open the survey and complete it the following day.

Cost

Prolific was cheaper based on the total cost compared with MTurk. Costs included the direct payment to participants, the base hosting fee paid to the platform, additional specifier fees, and taxes. A total of $1,155 was paid to MTurk, compared with $979 for Prolific, a difference of $176. The difference comes primarily from MTurk's "Premium Qualifications" fee, which cost $0.50 extra per participant to recruit only participants aged 18-25. The cost per high-quality participant was also calculated by dividing the total cost by the respective number of users who produced high-quality data (Douglas et al., 2023). Prolific ($6.40 per high-quality participant) was still cheaper than MTurk ($33 per high-quality participant).

Discussion

The aim of the current study was to compare the practical and data-driven differences between two popular participant pools, MTurk and Prolific, building upon work by Douglas et al. (2023), Webb and Tangney (2022), and others. Data analyzed in this study were drawn from a completed study conducted primarily to make conclusions about online dating behavior in young adults, with data collected across two crowdsourcing platforms: Amazon's Mechanical Turk and Prolific. This comparison sought to understand the cost and quality of data gathered across both platforms. Based on pricing and data quality–assessed through attention checks, duration, and internal consistency–Prolific proved to be the superior crowdsourcing platform compared to MTurk for these samples. Nonetheless, Prolific still demonstrated notable room for improvement within this sample.

In this study, only about a quarter of the sample produced high-quality data. Of the 188 that remained after screening, most (*n* = 153) came from Prolific, compared with MTurk (*n* = 35). Nearly 18% of MTurk participants fell outside the inclusionary age range–despite the added cost–an effect also observed by Webb and Tangney (2022). As a result, even the mean age (26.18) was outside of the inclusion criteria range (18-25). Internal consistency further supported Prolific; Cronbach's α values were higher for all factors except for Intellect/Imagination. Notably, the change in α values after screening demonstrates that participant exclusion based on data quality can alter conclusions, an idea supported by previous research (DeSimone & Harms, 2018). Finally, duration of the survey appeared equivalent for Prolific and MTurk.

Prolific also outperformed MTurk on cost. On an absolute basis, Prolific was cheaper ($979 USD) compared to MTurk ($1,155 USD) for gathering the same number of participants (*n* = 350). While compensation for the participants was held even ($2), the host fees and specifier charges led to the observed differences. A steep increase in cost may lead to a moral conundrum in which researchers may lower participant wages to afford the hosting of their survey. As a better alternative, we recommend opting for a cheaper platform, which depends on exclusion criteria (i.e., base rate and the need for specifiers). On a relative basis, Prolific was still the cheaper option. As determined through the cost per high-quality participant, MTurk participants necessitated $33 compared to $6.40 for Prolific participants. Effectively, we paid MTurk five times the U.S. dollar value for fewer "usable" participants. Based on this detailed comparison of the samples gathered, the authors perceive Prolific as the winner in this direct comparison between MTurk and Prolific.

Limitations & Future Directions

Several platform capabilities were not tested in the present study. This study was not longitudinal, so the tools that both companies offer for this type of research could not be compared as they have in other studies (e.g., Henderson et al., 2021; Kothe & Ling, 2019; Paas et al., 2018; Stoycheff, 2016). Additionally, this was an experimental psychology study that took around 15 minutes to complete. There is reason to believe that studies presented in different fields (e.g., Follmer et al., 2017; Reid et al., 2022; Wagner et al., 2021) and durations (e.g., Aguinis et al., 2021; Hamby & Taylor, 2016) may find different success with each platform.

Additionally, it is difficult to determine the source of low-quality data. It is quite common to read papers that describe the data spoilers as "bots" (e.g., Goodrich et al., 2023; Stokel-Walker, 2018; Webb & Tangney, 2022). However, deeper dives suggest that international participants, not "bots" or computer programs, are a primary source of lower data quality (Moss et al., 2021). International participants are often excluded, so they may lie about demographic information (e.g., native language and current location), which can confound results (Dennis et al., 2020). It is recommended that further research be conducted on these topics. Moreover, a reproduction of this study is warranted to evaluate ever-changing pools.

Recommendations for Researchers

As researchers develop increasingly sophisticated methods to detect low-quality data or robots, participants and programmers evolve strategies to evade detection. While there is no perfect solution, steps can be taken by researchers and crowdsourcing companies to improve the science generated on these platforms by filling their online surveys with relevant attention checks, participant verifiers, and logic.

Goodrich et al. (2023) recommend considering embedded survey components, including CAPTCHA, honeypot questions, and institutional knowledge checks to improve participant screeners. CAPTCHAs (Completely Automated Public Turing test to tell Computers and Humans Apart) vary in form, including checking the "I'm not a robot" box, selecting all of the pieces of the stoplight in a given picture, or decoding distorted letters (Moradi & Keyvanpour, 2015). Honeypot questions are ones embedded and hidden in surveys, making them invisible to human survey takers but visible to robots (Goodrich et al., 2023). If one is answered, you have good evidence that your survey was answered by an actual robot. Finally, institutional knowledge can be checked in a similar way to the aforementioned logic check implemented in the present study. Goodrich and colleagues (2023) suggest a question about the participant's zip code and then a follow-up about a nearby landmark, such as the closest university.

IP addresses can also be used to vet participants who have signed up more than once in one location (Aguinis et al., 2021). Unfortunately, several drawbacks are present when collecting IP addresses. Anonymity is violated, prohibiting a guarantee of identity protection. Moreover, even if identifiable data are secured, as they should be, this check would not guarantee that the participant is only completing the survey once. Most survey takers know that they can use free VPNs (Virtual Private Networks) that allow them to appear, to internet service providers, as if they are in different places across the world (Dennis et al., 2020). This also may unfairly disqualify multiple individuals who use the same device to participate, such as public library computers or devices shared between family members.

Aguinis and colleagues (2021) recommend considering response speed and consistency in the process of screening participants. Apart from reviewing the entire survey time, which should fall around a certain predetermined duration based on trials, researchers can look at individual question response times. It is unlikely that participants could respond to certain questions in under a second (Wood et al., 2017) unless they are extremely familiar with a given measure or the objective is to respond rapidly. Therefore, tracking question response time, can alert researchers to suspicious data. Moreover, inattentive participants can be identified if they mark the same response several times in a row (e.g., "*strongly agree*" for all ten questions on a given measure; Aguinis et al., 2021). Several methods exist to analyze response patterns of this sort that may be used to flag bots (DeSimone & Harms, 2018; Dunn et al., 2018).

Finally, researchers should become aware of techniques not implemented in this study or discussed herein to identify participants who supply low-quality data, lie about answers, or submit multiple responses. Several researchers have done excellent work in compiling recommendations, which should be reviewed in tandem with reflection on this paper (Aguinis et al., 2021; Goodrich et al., 2023; Hunt & Scheetz, 2019; Hydock, 2018; Kennedy et al., 2020; Newman et al., 2021; Sauter et al., 2020; Stanton et al., 2022).

Conclusion

This study leveraged a screening process similar to Webb and Tangney (2022), with heavy influence from Douglas et al. (2023), to compare MTurk and Prolific recruitment potential based on data quality and cost. Based on these metrics, Prolific outperformed MTurk for recruitment. However, while Prolific outperformed MTurk on our survey, researchers with different protocols may observe different results. Most meaningful-ly, researchers ought to critically evaluate the impact that using low-quality data in publications may have on societal outcomes for generations. As we found surprisingly few high-quality participants across both Prolific and MTurk, it is clear that improved survey methodologies are warranted regardless of platform. With this in mind, researchers should incorporate survey strategies demonstrated in this work as well as the highlighted best practices from other researchers.

References

Abbey, J. D., & Meloy, M. G. (2017). Attention by design: Using attention checks to detect inattentive respondents and improve data quality. *Journal of Operations Management, 53–56*, 63–70. https://doi.org/10.1016/j.jom.2017.06.001

Aguinis, H., Villamor, I., & Ramani, R. S. (2021). MTurk research: Review and recommendations. *Journal of Management, 47*(4), 823–837. https://doi.org/10.1177/0149206320969787

Albert, D. A., & Smilek, D. (2023). Comparing attentional disengagement between Prolific and MTurk samples. *Scientific Reports, 13*(1), 20574. https://doi.org/10.1038/s41598-023-46048-5

Amazon Mechanical Turk. (n.d.). Retrieved March 30, 2024, from https://www.mturk.com/

Bai, H. (2018). Evidence that a large amount of low quality responses on MTurk can be detected with repeated GPS coordinates. *Maxhuibai.com*. https://www.maxhuibai.com/blog/evidence-that-responses-from-repeating-gps-are-random

Barends, A. J., & de Vries, R. E. (2019). Noncompliant responding: Comparing exclusion criteria in MTurk personality research to improve data quality. *Personality and Individual Differences, 143*, 84–89. https://doi.org/10.1016/j.paid.2019.02.015

Best, S. J., Krueger, B., Hubbard, C., & Smith, A. (2001). An assessment of the generalizability of Internet surveys. *Social Science Computer Review, 19*(2), 131–145. https://doi.org/10.1177/089443930101900201

Burnette, C. B., Luzier, J. L., Bennett, B. L., Weisenmuller, C. M., Kerr, P., Martin, S., Keener, J., & Calderwood, L. (2022). Concerns and recommendations for using Amazon MTurk for eating disorder research. *International Journal of Eating Disorders, 55*(2), 263–272. https://doi.org/10.1002/eat.23614

Chandler, J., Paolacci, G., Peer, E., Mueller, P., & Ratliff, K. A. (2015). Using nonnaive participants can reduce effect sizes. *Psychological Science, 26*(7), 1131–1139. https://doi.org/10.1177/0956797615585115

Cheung, J. H., Burns, D. K., Sinclair, R. R., & Sliter, M. (2017). Amazon Mechanical Turk in organizational psychology: An evaluation and practical recommendations. *Journal of Business and Psychology, 32*(4), 347–361. https://doi.org/10.1007/s10869-016-9458-5

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*(1), 98–104. https://doi.org/10.1037/0021-9010.78.1.98

Crump, M. J. C., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PLOS ONE, 8*(3), e57410. https://doi.org/10.1371/journal.pone.0057410

Dennis, S. A., Goodson, B. M., & Pearson, C. A. (2020). Online worker fraud and evolving threats to the integrity of MTurk data: A discussion of virtual private servers and the limitations of IP-based screening procedures. *Behavioral Research in Accounting, 32*(1), 119–134. https://doi.org/10.2308/bria-18-044

DeSimone, J. A., & Harms, P. D. (2018). Dirty data: The effects of screening respondents who provide low-quality data in survey research. *Journal of Business and Psychology, 33*(5), 559–577. https://doi.org/10.1007/s10869-017-9514-9

Donnellan, M., Oswald, F., Baird, B., & Lucas, R. (2006). The Mini-IPIP scales: Tiny-yet-effective measures of the Big Five factors of personality. *Psychological Assessment, 18*, 192–203. https://doi.org/10.1037/1040-3590.18.2.192

Douglas, B. D., Ewell, P. J., & Brauer, M. (2023). Data quality in online human-subjects research: Comparisons between MTurk, Prolific, CloudResearch, Qualtrics, and SONA. *PLOS ONE, 18*(3), e0279720. https://doi.org/10.1371/journal.pone.0279720

Dunn, A. M., Heggestad, E. D., Shanock, L. R., & Theilgard, N. (2018). Intra-individual response variability as an indicator of insufficient effort responding: Comparison to other indicators and relationships with individual differences. *Journal of Business and Psychology, 33*(1), 105–121. https://doi.org/10.1007/s10869-016-9479-0

Follmer, D. J., Sperling, R. A., & Suen, H. K. (2017). The role of MTurk in education research: advantages, issues, and future directions. *Educational Researcher, 46*(6), 329–334. https://doi.

org/10.3102/0013189X17725519

Fong, D. Y., Ho, S. Y., & Lam, T. H. (2010). Evaluation of internal reliability in the presence of inconsistent responses. *Health and Quality of Life Outcomes, 8*, 27. https://doi.org/10.1186/1477-7525-8-27

Goodrich, B., Fenton, M., Penn, J., Bovay, J., & Mountain, T. (2023). Battling bots: Experiences and strategies to mitigate fraudulent responses in online surveys. Applied Economic *Perspectives and Policy, 45*(2), 762–784. https://doi.org/10.1002/aepp.13353

Gosling, S. D., & Mason, W. (2015). Internet research in psychology. *Annual Review of Psychology, 66*, 877–902. https://doi.org/10.1146/annurev-psych-010814-015321

Graneheim, U. H., Lindgren, B. M., & Lundman, B. (2017). Methodological challenges in qualitative content analysis: A discussion paper. *Nurse Education Today, 56*, 29-34. https://doi.org/10.1016/j.nedt.2017.06.002

Hamby, T., & Taylor, W. (2016). Survey satisficing inflates reliability and validity measures: An experimental comparison of college and Amazon Mechanical Turk samples. *Educational and Psychological Measurement, 76*(6), 912–932. https://doi.org/10.1177/0013164415627349

Hauser, D. J., Ellsworth, P. C., & Gonzalez, R. (2018). Are manipulation checks necessary? *Frontiers in Psychology, 9*. https://doi.org/10.3389/fpsyg.2018.00998

Hauser, D. J., & Schwarz, N. (2016). Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior Research Methods, 48*(1), 400–407. https://doi.org/10.3758/s13428-015-0578-z

Henderson, E. L., Simons, D. J., & Barr, D. J. (2021). The trajectory of truth: A longitudinal study of the illusory truth effect. *Journal of Cognition, 4*(1), 29. https://doi.org/10.5334/joc.161

Hunt, N. C., & Scheetz, A. M. (2019). Using MTurk to distribute a survey or experiment: methodological considerations. *Journal of Information Systems, 33*(1), 43–65. https://doi.org/10.2308/isys-52021

Hydock, C. (2018). Assessing and overcoming participant dishonesty in online data collection. *Behavior Research Methods, 50*(4), 1563–1567. https://doi.org/10.3758/s13428-017-0984-5

Kemp, S. (2023, January 26). *Digital 2023: Global overview report.* DataReportal – Global Digital Insights. https://datareportal.com/reports/digital-2023-global-overview-report

Kennedy, R., Clifford, S., Burleigh, T., Waggoner, P. D., Jewell, R., & Winter, N. J. G. (2020). The shape of and solutions to the MTurk quality crisis. *Political Science Research and Methods, 8*(4), 614–629. https://doi.org/10.1017/psrm.2020.6

Kothe, E. J., & Ling, M. (2019, September 6). Retention of participants recruited to a multi-year longitudinal study via Prolific. *PsyArXiv.* https://doi.org/10.31234/osf.io/5yv2u

Litman, L., Moss, A., Rosenzweig, C., & Robinson, J. (2021). Reply to MTurk, Prolific or panels? Choosing the right audience for online research. *SSRN*. https://doi.org/10.2139/ssrn.3775075

Litman, L., Robinson, J., & Rosenzweig, C. (2015). The relationship between motivation, monetary compensation, and data quality among US- and India-based workers on Mechanical Turk. *Behavior Research Methods, 47*(2), 519–528. https://doi.org/10.3758/s13428-014-0483-x

Lu, L., Neale, N., Line, N. D., & Bonn, M. (2022). Improving data quality using Amazon Mechanical Turk through platform setup. *Cornell Hospitality Quarterly, 63*(2), 231–246. https://doi.org/10.1177/19389655211025475

Matjašič, M., Vehovar, V., & Manfreda, K. L. (2018). Web survey paradata on response time outliers: A systematic literature review. *Advances in Methodology and Statistics, 15*(1), 23-41. https://ibmi.mf.uni-lj.si/mz/2018/no-1/Matjasic2018.pdf

Matthijsse, S. M., de Leeuw, E. D., & Hox, J. J. (2015). Internet panels, professional respondents, and data quality. *Methodology, 11*(3), 81–88. https://doi.org/10.1027/1614-2241/a000094

Moradi, M., & Keyvanpour, M. (2015). CAPTCHA and its alternatives: A review. *Security and Communication Networks, 8*(12), 2135–2156. https://doi.org/10.1002/sec.1157

Moss, A. J., Rosenzweig, C., Jaffe, S. N., Gautam, R., Robinson, J., & Litman, L. (2021). Bots or inattentive humans? Identifying sources of low-quality data in online platforms. *PsyArXiv.* https://doi.org/10.31234/osf.io/wr8ds

Newman, A., Bavik, Y. L., Mount, M., & Shao, B. (2021). Data collection via online platforms: challenges and recommendations for future research. *Applied Psychology, 70*(3), 1380–1402. https://doi.org/10.1111/apps.12302

Paas, L. J., Dolnicar, S., & Karlsson, L. (2018). Instructional manipulation checks: A longitudinal analysis with implications for MTurk. *International Journal of Research in Marketing, 35*(2), 258–269. https://doi.org/10.1016/j.ijresmar.2018.01.003

Palan, S., & Schitter, C. (2018). Prolific.ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance, 17*, 22–27. https://doi.org/10.1016/j.jbef.2017.12.004

Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology, 70*, 153–163. https://doi.org/10.1016/j.jesp.2017.01.006

Peer, E., Rothschild, D., Gordon, A., Evernden, Z., & Damer, E. (2022). Data quality of platforms and panels for online behavioral research. *Behavior Research Methods, 54*(4), 1643–1662. https://doi.org/10.3758/s13428-021-01694-3

Prolific (n.d.). Retrieved March 30, 2024, from https://www.prolific.com/

Reid, B., Wagner, M., d'Amorim, M., & Treude, C. (2022). Software engineering user study recruitment on Prolific: An experience report. *ArXiv*. https://doi.org/10.48550/arXiv.2201.05348

Sauter, M., Draschkow, D., & Mack, W. (2020). Building, hosting and recruiting: A brief introduction to running behavioral experiments online. *Brain Sciences, 10*(4), 251. https://doi.org/10.3390/brainsci10040251

Simone, M. (2019, November 21). Bots started sabotaging my online research. I fought back. *STAT.* https://www.statnews.com/2019/11/21/bots-started-sabotaging-my-online-research-i-fought-back/

Smith, N. A., Sabat, I. E., Martinez, L. R., Weaver, K., & Xu, S. (2015). A convenient solution: using MTurk to sample from hard-to-reach populations. *Industrial and Organizational Psychology, 8*(2), 220–228. https://doi.org/10.1017/iop.2015.29

Stanton, K., Carpenter, R. W., Nance, M., Sturgeon, T., & Villalongo Andino, M. (2022). A multisample demonstration of using the prolific platform for repeated assessment and psychometric substance use research. *Experimental and Clinical Psychopharmacology, 30*(4), 432–443. https://doi.org/10.1037/pha0000545

Stokel-Walker, C. (2018, August 10). Bots on Amazon's Mechanical Turk are ruining psychology studies. *New Scientist.* https://www.newscientist.com/article/2176436-bots-on-amazons-mechanical-turk-are-ruining-psychology-studies/

Stoycheff, E. (2016). Please participate in Part 2: Maximizing response rates in longitudinal MTurk designs. *Methodological Innovations, 9.* https://doi.org/10.1177/2059799116672879

Teitcher, J. E. F., Bockting, W. O., Bauermeister, J. A., Hoefer, C. J., Miner, M. H., & Klitzman, R. L. (2015). Detecting, preventing, and responding to "fraudsters" in Internet research: Ethics and tradeoffs. *Journal of Law, Medicine & Ethics, 43*(1), 116–133. https://doi.org/10.1111/jlme.12200

Wagner, A., Bakas, A., Kennison, S., & Chan-Tin, E. (2021). A comparison of SONA and MTurk for cybersecurity surveys. *Proceedings of the 2021 European Interdisciplinary Cybersecurity Conference*, 87–88. https://doi.org/10.1145/3487405.3487657

Webb, M. A., & Tangney, J. P. (2022). Too good to be true: Bots and bad data from Mechanical Turk. *Perspectives on Psychological Science. 19*(6), 887-890. https://doi.org/10.1177/17456916221120027

Wood, D., Harms, P. D., Lowman, G. H., & DeSimone, J. A. (2017). Response speed and response consistency as mutually validating indicators of data quality in online samples. *Social Psychological and Personality Science, 8*(4), 454–464. https://doi.org/10.1177/1948550617703168

Zack, E. S., Kennedy, J., & Long, J. S. (2019). Can nonprobability samples be used for social science research? A cautionary tale. *Survey Research Methods, 13*(2), 215-227. https://doi.org/10.18148/srm/2019.v13i2.7262

Table 1

Summary of Results from the Screening Process

Screener Steps	MTurk (*n* = 354)	Prolific (*n* = 345)	Total (*n* = 699)
Age	229 (64.69%)	345 (100%)	574 (82.12%)
Seriousness	227 (64.12%)	344 (99.71%)	571 (81.69%)
Open-ended	64 (18.08%)	229 (66.38%)	293 (41.92%)
Other Sensible	35 (9.89%)	153 (44.35%)	188 (26.90%)
	Cost		
Total Cost	$1,155	$979	$2,134
Cost per high-quality participant	$33	$6.40	$11.35

Table 2

Reliability Metrics for the Validated Mini-IPIP (Donnellan et al., 2006), MTurk, and Prolific

	Mini-IPIP				MTurk			Prolific		
	α	Mean	*SD*		α	Mean	*SD*	α	Mean	*SD*
Extraversion	.77	3.28	.90	Before	.35	2.94	.76	.42	2.82	.26
				After	.80	2.67	.21	.86	2.72	.26
Agreeableness	.70	4.01	.69	Before	.29	3.26	.82	.51	4.02	.23
				After	.72	3.74	.36	.78	3.99	.21
Conscientiousness	.69	3.42	.78	Before	.26	3.13	.79	.38	3.58	.30
				After	.38	3.58	.24	.77	3.56	.34
Neuroticism	.68	2.54	.80	Before	.002	2.88	.83	.41	2.95	.38
				After	.57	2.78	.37	.78	2.91	.39
Intellect/ Imagination	.65	3.70	.73	Before	.52	2.98	.64	.37	4.00	.12
				After	.83	3.52	.05	.73	3.99	.12

Note. “Before” signifies the data prior to screening, and “After” signifies the data following screening.

Graduate Student Journal of Psychology
Fall 2025 - Vol. 25

Response Time to Detect Careless Responding and Its Relationship with and Prediction of Emotional Distress

Kristen E. Zentner & Seyma N. Yildirim-Erbasli
Department of Psychology, Concordia University of Edmonton

People experiencing emotional distress struggle with cognitive and motivational decline, which has been correlated with patterns of careless responding. Although several methods have been used to detect careless responses in emotionally distressed respondents, the response time has not been widely explored. The current study conducted secondary data analyses on a sample (N = 37,819) who completed the Depression Anxiety Stress Scale (DASS-42) in an online survey between 2017 and 2019. First, a response-time-based approach––a normative threshold method––was used to identify careless responding and examine its association with emotional distress using the DASS-42. Second, four machine learning models––decision tree (DT), random forest (RF), support vector machine (SVM), and naive Bayes (NB)––were trained on DASS-42 item responses and response times to predict emotional distress severity level. A significant correlation was found between the number of careless responses and subscale scores of anxiety and stress. In addition, Mann-Whitney U tests showed statistically significant differences between careless and careful responders in depression, anxiety, and stress. Regarding the machine learning models, SVM was found to be the best predictive model for classifying distressed people with an accuracy, sensitivity, and specificity exceeding 90%. Our results suggest that, in addition to survey responses, response time can identify careless responders and predict distressed responders.

Keywords: Response time, machine learning, psychological distress, careless responding

Computerized self-report measures have revolutionized the administration of psychological measures by offering researchers and clinicians a more efficient and convenient means of collecting data. However, the potential of computerized testing has yet to be fully realized. Additional data available in computerized testing (e.g., response time), which are not available through traditional pen-and-paper administration, offer significant contributions to the psychometric utility of these tests. Response time is recognized for its ability to provide a broader representation of responses, going beyond the responses themselves (van der Linden et al., 2010). It has considerable potential to reveal psychometrically relevant information, assess profile validity (i.e., the extent to which an individual's test score represents the true level of the trait or ability being measured), and develop increasingly precise and accurate adaptive testing methods.

Emotional Distress and Response Behavior

There is strong evidence to suggest the impact of emotion on performance on cognitive tasks in the cognitive psychology literature (e.g., Castaneda et al., 2008, 2011; Eysenck et al., 2007; Gross, 2015; Hubbard et al., 2016). Given that emotion and cognition interact to impact behavior, it is necessary for experimental researchers to consider how this interaction may impact the quality of self-report data in various contexts. For example, research indicates that individuals with greater levels of emotional distress have biases in their self-report data linked to this level of emotionality (Ashley & Shaughnessy, 2021; Conijn et al., 2020). Furthermore, researchers are often interested in measuring emotion itself. The Profile of Mood States (POMS; McNair et al., 1971), the Multiple Affect Adjective Checklist (MAACL; Zuckerman & Lubin, 1985), the Positive and Negative Affect Scale (PANAS-X; Watson & Clark, 1994), and the Depression Anxiety and Stress Scale (DASS-42; Antony et al., 1998) are frequently used self-reports that measure various aspects of emotional states. When a participant completes such self-reports, they are engaging in a cognitive task (i.e., completing the survey itself) that involves emotional content (i.e., the item content) which may interact with or influence the emotional or cognitive state of the participant, such as instigating a heightened emotional response or engagement in emotional regulation strategies (Castaneda et al., 2008; Gotlib & Joormann, 2010; Gross, 2015; Hubbard et al., 2016; Sun & Alkon, 2014). Despite recent research, gaps exist in our understanding of how emotion and cognition interact to impact response behavior on surveys.

Cognitive symptoms of emotional distress can impact data quality on self-report measures. Self-report surveys require a degree of effort to complete, and factors that compromise sustained effort may impact survey response styles. Several theoretical models have been applied to understand the relationship between emotion and cognition in response behavior. For example, cognitive bias literature suggests that mood-congruent information increases working memory in depressed individuals; that is, they have a bias to pay attention to information that reaffirms their depressive cognitions (Hubbard et al., 2016). However, drawing on cognitive behavioral theory, Ashley and Shaughnessy (2021) proposed that individuals with depression engage in avoidance behaviors when completing surveys, adopting strategies that reduce the impact of distressing survey content by minimizing effort, attention, and time to completion. This calls into question whether self-reports of emotional distress are responded to with heightened attention due to the mood-congruent nature of the items, or with inattention (i.e., careless responding) due to concentration difficulties and avoidance behaviors characteristic of depression, anxiety, and stress symptoms. It is recognized that the cognitive impairments associated with depression, anxiety, and stress impact how these individuals respond to self-report measures, potentially threatening their profile validity and rendering scores inaccurate (Ashley & Shaughnessy, 2021; Conijn et al., 2020). Studying the patterns of response behavior on self-report measures in emotionally distressed respondents may clarify how their cognitive symptoms are impacting their survey responses and provide an effective way to assess profile validity.

Careless responding tends to be higher in populations with mental health concerns, with rates ranging from 6.0% (LePagea et al., 2001) to 12.6% (Conijn et al., 2015). Moreover, findings indicate that those with more severe psychopathology are more likely to show aberrant response styles (Conijn et al., 2015, 2018; Keeley et al., 2016; Wardenaar et al., 2015). Comorbid anxiety and depression appear to be associated with even greater aberrant responses because of the interacting effects of the two forms of psychological distress on working memory capacity (e.g., Beaudreau & O'Hara, 2009). The relationship between anxiety alone and response bias is less clear (Ferreri et al, 2011; Salthouse, 2012), possibly because those with anxiety do not consistently exhibit cognitive symptoms (Castaneda et al., 2008). Conijn et al. (2020) proposed and tested a theoretical explanation for careless responding in clinically depressed and anxious populations. They argued that cognitive symptoms of depression, specifically concentration, comprehension, and memory, limit cognitive abilities and make aberrant responses more likely (Hubbard et al., 2016). In testing their model, they found that cognitive symptoms mediate the relationship between depression or anxiety and response biases. Another study indicated that higher levels of anxiety, distress, and sadness were associated with inattention on surveys (Ashley & Shaughnessy, 2021).

When researchers fail to detect and report instances of careless responding, it impacts findings, jeopardizing the overall quality of knowledge production in the field. This threat to the psychometric properties of self-report measures has been reported and studied by many researchers, and its relevance to clinical settings has been explored (e.g., Cuijpers et al., 2010; Keeley et al., 2016; Tada et al., 2014). When self-report questionnaires are used diagnostically, clinicians base their clinical decision-making and diagnosis on information that may overestimate or underestimate symptom severity (Keeley et al., 2016). Given the findings from previous research and the psychometric utility of attending to careless responding, a clear understanding of careless response detection is needed.

Approaches to Understanding Response Behavior

There are a variety of approaches to detect careless responding (see Ward & Meade, 2023, for review). Proactive indices are those that place items within the survey itself to assess inattention, such as "Because I am paying attention, I will answer this question with '*Very little*'" (Ashley & Shaughnessy, 2021, p. 4). However, these single-item proactive indices provide little contextual information about the pattern of careless responding throughout an entire survey. Reactive indices are those that flag inattentive responders through detection of careless response styles during data cleaning and analysis phases, such as longstring detection (i.e., the selection of the same response option for several consecutive responses), and participant-specific reliability (i.e., the consistency of a participant's responses on items measuring the same trait; Ashley & Shaughnessy, 2021). Curran (2016) noted that the use of a single response style approach is insufficient to detect careless responders because they may have response style patterns that are detectable with some approaches but not others. For example, participants who use a longstring response style would have high participant-specific reliability and would not be identified by detection approaches designed to detect even-odd response styles where participants select extreme ends of a scale (Meade & Craig, 2012). To account for this diversity in response styles, researchers have suggested that the use of multiple detection approaches is necessary to identify careless responders (Ashley & Shaughnessy, 2021; Curran, 2016).

In their review of careless responding, Ward and Meade (2023) suggest that extensive screening methods may be necessary when analyzing large datasets or with populations that are more likely to engage in careless responding, such as emotionally distressed individuals. For example, Ashley and Shaughnessy (2021) found that proactive items and short survey response time were associated with negative emotional states (i.e., sadness, anxiety, distress) while other detection methods (e.g., longstring, participant-specific reliability) were not.

Response Time Approaches

Response time approaches, which operate on the assumption that careless responders will have unreasonably rapid survey response times consistent with the motivation to finish the survey quickly, have received considerable review within the literature (Ashley & Shaughnessy, 2021; Curran, 2016; Jones et al., 2022; Ward & Meade, 2022). Researchers have identified that response time has the potential as a detection method on computerized tests because it is more difficult to manipulate than other methods (e.g., longstring, participant-specific reliability; Curran, 2016). For example, careless responders who wish to appear as careful may be motivated to provide response patterns that mimic normal responses (e.g., selecting responses consistently at one end of the scale with enough variability to avoid pattern detection of longstring and even-odd styles). Due to their desire to finish the survey quickly, however, they will likely still have shorter response times than careful responders, as demonstrated by Schnipke and Scrams (1997).

Careless and careful responders exhibit distinct distributions in response time, with an initial spike in response time distribution attributed to the former. Curran (2016) used simulated data to explore distributions of response time in careless and careful responders and highlighted consideration of Type I (i.e., falsely identifying a careful responder as careless) and Type II errors (i.e., falsely identifying a careless responder as careful) in establishing cut-off scores due to the significant overlap between distributions. Researchers have used various calculations to determine cut-off scores, including 1.5 quartiles above or below the median (Funke, 2016), two standard deviations above or below the mean (Heerwegh, 2003), one standard deviation above or below the mean (Ashley & Shaughnessy, 2021), and various percentiles (e.g., first percentile, fifth percentile; Gummer & Roßmann, 2015; Harms et al., 2017). Some researchers have explored cut-off scores on an individual item level. For example, a 2-second-per-item cut-off score is considered a conservative approach, limiting Type II errors at the cost of missing some careless responders (Bowling et al., 2016; Huang et al., 2012).

Several methods have been used to determine response time thresholds for individual items, including a two-state mixture model (Schnipke & Scrams, 1997), surface-level characteristics of items (i.e., character count; Wise & Kong, 2005), and visual inspection of the response time-frequency distribution (Wise, 2006). These three approaches tend to identify similar item response time thresholds (Kong et al., 2007). Considering that items vary in the amount of text or how mentally taxing they are, what may be classified as rapid responding varies by the item. This implies that greater accuracy of careless response detection may be gained by identifying a normative threshold specific to each item, rather than using a general response time threshold applied to all questions as has been done previously (Huang et al., 2012; Wise & Ma, 2012). The normative threshold approach involves the use of response time cut-off scores to identify individual items that are responded to carelessly (Wise & Ma, 2012).

Response time approaches have been of limited utility in psychological research compared to other careless responding detection methods because this metric is typically available at a page or survey level (Ashley & Shaughnessy, 2021; Ward & Meade, 2023). However, with item response time available at the item level, researchers and clinicians can gain precision in identifying response time patterns of careless responders in emotional distress. The greatest contribution of item response time to the careless responding literature might be its potential to detect careless responders regardless of the responders' specific response style (e.g., long string), which gives researchers detailed information on survey response patterns while blocking respondents' attempts to mask their response style (Curran, 2016).

Computerized surveys that provide access to response time at the item level also create opportunities for more sophisticated analyses. For example, computerized adaptive testing uses an algorithm to select items based on previous responses to gain efficiency and precision in the measurement of the ability or trait with the administration of fewer items (Wise, 2020). Companies that produce widely used psychological measures are increasingly moving to computerized adaptive formats for reduced testing time and ease of administration and scoring (Forbey et al., 2012). Wise (2020) notes that traditional approaches to computerized adaptive testing in education, which use only item difficulty in adaptation algorithms, could be expanded to use behavioral measures such as response time to improve the precision and accuracy of measurement. Computerized adaptive testing relies on advanced modeling, such as machine learning approaches, which can be used to examine response and response time patterns to predict emotional distress.

Machine Learning Approaches to Predict Emotional Distress

Several researchers have used machine learning methods with item response data to explore patterns that inform psychologists about how emotionally distressed individuals respond to self-report surveys. This line of research stems from machine learning's ability to capture subtle patterns not evident through traditional approaches and its effectiveness in handling the complex interactions and dependencies among variables, which are likely to be common in psychological assessment data collected by surveys such as the DASS. For example, Budiyanto et al. (2019) used text mining from social media posts to measure depression and anxiety using a closed-loop machine learning approach, with an NB algorithm as a training process and the DASS-21 parameters as a learning process. From this approach, Budiyanto et al. (2019) demonstrated the usefulness of machine learning methods to collect novel information about emotional distress. Other studies have found that various machine learning models are useful in predicting depression, anxiety, and stress from the DASS-42 and DASS-21, with some machine learning models showing greater accuracy and efficiency than others (Kumar et al., 2020; Priya et al., 2020; Srinath et al., 2022).

Kumar et al. (2020) predicted five severity levels of emotional distress by investigating eight machine learning algorithms trained on DASS-42 item responses, and then the same methods were applied to a second DASS-21 dataset. The results showed that these models could be used to predict emotional distress, with accuracy rates between 96.02% and 97.48% for the subscales (Kumar et al., 2020). Priya et al. (2020) applied five machine learning models trained on item response data, including DT, RF, NB, SVM, and *K*-nearest neighbor, to predict depression, anxiety, and stress levels from a sample of DASS-21 data. They found that the RF classifier demonstrated the best performance, with accuracy rates between 71.4% and 79.8% for DASS subscales. Srinath et al. (2022) compared SVM and logistic regression using parameter tuning to predict depression, anxiety, and stress from DASS-42 item response data. They found that logistic regression had the highest performance, with an accuracy of 98.15% for depression, 98.05% for anxiety, and 98.45% for stress. In recent years, researchers have begun to explore the utility of machine learning models using response time on behavioral tasks (i.e., perceptual matching task) and using neuroimaging and physiological data (i.e., Magnetic Resonance Imaging; Liu et al., 2022). These studies suggest that machine learning has predictive potential within clinical and counseling psychology. Moreover, Priya et al. (2020) noted that the sensitivity and specificity afforded by machine learning models make these approaches particularly helpful within healthcare contexts.

Current Research

Considering how often self-report is used to measure, research, and reduce impairment from negative emotional states such as depression, anxiety, and stress, further exploration of careless response identification is needed to enhance data quality and continue to elucidate the impact of emotion on cognitive tasks. While response time has been used as a measure of response behavior (e.g., Kong et al., 2007), more study of careless response detection within emotionally distressed populations is needed to determine effective ways to identify and deal with potentially invalid data. In addition, the use of machine learning approaches can facilitate the identification of complex patterns and relationships that may not be apparent through traditional statistical methods. There is a recent body of literature on the use of machine learning approaches to predict emotional distress using item response data (e.g., Srinath et al., 2022). However, this study proposes incorporating response time in addition to item responses, as response time can provide insights beyond the responses themselves (van der Linden et al., 2010). The present exploratory study aims to address these gaps in the literature by examining the relationship between careless responding and emotional distress, as well as exploring the potential utility of incorporating response time in machine learning models for predicting emotional distress. In this paper, the following research questions were investigated:

1. Is there an association between careless responding and emotional distress (i.e., depression, anxiety, and stress)?
2. Can machine learning models identify emotionally distressed people using item responses and response time?

Methods

The DASS-42 is a well-established measure of emotional distress with 42 items such as, “I felt that life was meaningless” in the depression subscale, “I was aware of dryness in my mouth” in the anxiety subscale, and “I found that I was very irritable” in the stress subscale on a scale of 0 (Did not apply to me at all) to 3 (Applied to me very much or most of the time; Lovibond & Lovibond, 1995). The subscales assess depression ($\alpha = .97$), anxiety ($\alpha = .92$), and stress ($\alpha = .95$) as separate constructs with 14 items each, and each demonstrates high internal consistency (Antony et al., 1998).

To address the research questions in this study, the dataset was pulled from the Open Source Psychometrics Project (2019), which offers public datasets. The survey was open for anyone to complete, meaning the sample may include a mixture of clinical and non-clinical populations. Participants received only their personalized results in return for their participation. The dataset was pre-cleaned upon download––negative response times were recoded to missing values, and milliseconds were transformed into seconds. Data analysis was conducted using jamovi and R programming languages. The sample consisted of 37,819 participants who completed the online survey on a scale of 1 to 4 between 2017 and 2019, aged 13 to 79 years (M = 23.39, SD = 8.57). Sociodemographic characteristics of the sample are displayed in Table 1.

Question 1: Association between Careless Responding and Emotional Distress

Figure 1 encapsulates the methodology employed to address our primary research question. To answer our first research question, we identified careless responses by calculating a normative threshold based on Wise and Ma's (2012) response time approach. Second, we classified participants as careless or careful upon considering the total number of careless responses they exhibited. Third, we examined the association between the number of careless responses and emotional distress scores. Finally, we conducted Mann-Whitney U tests to compare emotional distress scores between careless and careful responders.

The normative threshold for an item is calculated as "a percentage of the elapsed time between when the item is displayed and the mean of the response time distribution for the item, up to a maximum threshold value of ten seconds" (Wise & Ma, 2012, p. 9). For example, if an item takes a mean of 60 seconds for participants to complete, the 10% normative threshold (i.e., NT10) would be six seconds. Various normative thresholds can be compared to determine the cut-off that yields the greatest accuracy for identifying careless responders (Wise & Ma, 2012). We chose to use the 20% normative threshold (NT20) based on literature showing that this cut-off is appropriate for low-stakes environments (Rios & Soland, 2021) and due to the positively skewed distribution of DASS-42 scores in the dataset. To calculate the NT20, the mean response time was calculated for each item, and 20% of the average response time served as the NT20 cut-off score. We recoded careless responses (i.e., responses below the NT20 cut-off) as missing values. The reason behind this is that previous studies have pointed out how the presence of careless responses in the dataset can introduce bias into estimates of item and person parameters (e.g., Guo et al., 2016).

Various cut scores have been used to classify careful and careless responders (i.e., demonstrating a substantial number of careless responses throughout the survey; e.g., Wise & Kong, 2005). The purpose of the tool and the sample under study are crucial factors in determining appropriate cut scores. For example, a cut score of 20% was used in a low-stakes assessment (e.g., Wise & Kong, 2005). When a survey is used diagnostically for clinical purposes, prudent clinicians must be confident that the data are not impacted by careless responding, while putting more weight on other data sources (i.e., interviews) if the survey data have questionable validity (American Psychological Association, 2020). Stated another way, increasing false positives (classifying careful responders as careless) may be necessary for evaluating the validity of clinically relevant data. Given that the survey in the current study measures clinically relevant variables, a cut score of 10% was chosen. By setting the threshold at this level, we aimed to be inclusive enough to detect individuals who may exhibit a notable pattern of inattentive responses across the survey items. At the same time, the 10% threshold is chosen to avoid categorizing individuals as careless responders when they may, in fact, be providing thoughtful and considered answers to the survey questions. Participants who showed between 0 and 4 careless responses within the DASS-42 items were classified as careful responders (n = 37,025), and participants who showed between 5 and 20 careless responses were classified as careless responders (n = 551). Participants with 20 or more careless responses were considered extremely careless responders (n = 243) and were excluded from analyses because their survey scores would have been significantly biased by the severity of their careless responding. This is consistent with previous literature that uses a 50% careless response rate as a cut-off for removal from the dataset (Arias et al., 2020; Curran, 2016).

Given that data from carelessly responded items is invalid and introduces bias (e.g., Guo et al., 2016), it is necessary to recode these responses as missing values and compute scores accordingly. Therefore, after identifying careless responses, DASS-42 subscale scores were adjusted to represent only carefully responded items by recoding these responses as missing values and calculating the adjusted total scores. The adjusted subscale score is the sum of the scores for carefully responded items divided by the maximum possible total score for those items. We presented the adjusted scores as percentages to facilitate the interpretation of the results. For example, if a participant had careless responses on three items on the depression subscale (14 items), their depression subscale score would be the sum of their scores on the remaining carefully responded 11 items, with a score range of 11 (11 × 1 point) to 44 (11 × 4 points). If the participant scored 39 on these 11 carefully responded depression items, their adjusted subscale score in percentage would be calculated as follows: $39/44 = 0.87 \times 100 = 87$.

A Spearman correlation analysis was used to determine whether an association exists between the number of careless responses and emotional distress scores. Next, Mann-Whitney U tests were used to compare DASS-42 subscale scores between participants classified as careless and careful.

Question 2: Predicting Emotional Distress with Ma-

chine Learning

Figure 2 encapsulates the methodology employed to address our second research question. To answer our second research question, we used DASS-42 data consisting of item responses and response times of 37,819 respondents. Responders were classified into five severity levels on an ordinal scale for depression, anxiety, and stress based on their scores using the guide for severity levels of emotional distress in DASS-42 (see Table 2). Total scores were calculated on a 0 to 3 scale by subtracting one from each response, as the initial dataset included responses ranging from 1 to 4. For each emotional distress, we used 80% of the dataset for training and 20% for testing. Validation was conducted within the training process through 10-fold cross-validation (i.e., the number of groups that the dataset is randomly split into) and a random search for hyperparameter optimization.

We trained four machine learning algorithms–DT, RF, SVM, and NB (James et al., 2017)–to predict the severity levels of responders for each emotional distress. DTs split data into progressively smaller subsets based on selected features to form a simple, interpretable tree structure. RFs enhance this approach by combining multiple DTs built from random subsets of data and features to improve accuracy and reduce overfitting. NB applies Bayes' theorem under the assumption that all features are independent to generate probability-based classifications. SVMs identify the optimal hyperplane with the maximum margin to separate data points for predictive performance. After the training, we evaluated the performance of the trained models using the test sets and reported standard classification metrics (i.e., sensitivity, specificity, and accuracy).

Results

Careless Responding and Emotional Distress

There were statistically significant correlations between the number of careless responses and subscales of anxiety ($r_s(37{,}574) = .03$, $p < .001$) and stress ($r_s(37{,}574) = .02$, $p = .001$), but not the subscale of depression ($r_s(37{,}574) = -.001$, $p = .836$).

In terms of the subscale of depression, careless responders (*Mdn* = 67.3) had higher scores than careful responders (*Mdn* = 62.5), and the difference between careless and careful responders was statistically significant ($U = 9.58e{+}6$, $p = .01$, $r = .06$; see Figure 3). Similarly, a statistically significant difference was found between groups on the anxiety subscale ($U = 9.00e{+}6$, $p < .001$, $r = 0.12$) with careless responders (*Mdn* = 57.1) scoring higher on anxiety than careful responders (*Mdn* = 51.8). Finally, higher stress subscale scores were found among careless responders (*Mdn* = 65.9) than careful responders (*Mdn* = 62.5) with a statistically significant difference ($U = 9.62e{+}6$, $p = .02$, $r = .06$). In summary, careless responders had higher scores of depression, anxiety, and stress than careful responders.

Predicting Emotional Distress

Table 3 shows single classification metrics of the four machine learning models for five different severity levels of each emotional distress. Even though machine learning models showed roughly similar performance, DT, RF, and NB yielded inconsistent sensitivity values for mild, moderate, and severe levels. In terms of the depression subscale, classification metrics for almost every level exceeded the acceptable threshold of 70% or the optimal threshold of 80% with several exceptions of sensitivity values being less than the acceptable threshold. Overall, the SVM outperformed the other models with classification metrics exceeding the optimal threshold for each level. Regarding the anxiety subscale (see Table 3), with the exception of SVM, the other models showed mixed and extremely low sensitivity values for the levels of mild, moderate, and severe, while others exceeded the optimal threshold of 80%. Similar to depression, SVM surpassed the other models with classification metrics above the optimal threshold for each level. For the stress subscale (see Table 3), DT, RF, and NB showed mixed and low sensitivity values for the levels of mild, moderate, and severe. Other measures went over the optimal threshold of 80% and, in particular, SVMs dominated the other models in terms of sensitivity and specificity, which were either 100% or very close.

We also macro-averaged the single metrics by calculating the averages of sensitivity and specificity values over severity levels (see Figure 4). Based on the macro-averaging, four machine learning models for three types of emotional distress showed similar results in terms of specificity values being larger than 90%, exceeding the optimal threshold, whereas they showed mixed results for sensitivity. Only SVM exceeded the optimal threshold, with sensitivity values being larger than 90% across all three types of emotional distress. These results reveal that true negatives can be predicted with optimal sensitivity and specificity by all four classification models. However, both true positives and true negatives can be predicted with optimal sensitivity and specificity by only SVM. In addition, among all four machine learning methods, SVM had the highest accuracy of classification compared to the other methods across all three emotional distress, followed by RF, DT, and NB (see Figure 5).

Discussion

The current exploratory study contributes to the literature by investigating the relationship between careless responding and self-reported depression, anxiety, and stress using response behavior information (i.e., response time). Additionally, the inclusion of item response time in machine learning models was explored for predicting depression, anxiety, and stress severity levels. These findings are germane to researchers who use computerized self-report measures, as response time can aid in the identification of careless responders who bias datasets and invalidate individual testing profiles. Furthermore, machine learning models can efficiently predict the severity level of emotional distress by taking not only item responses but also response time patterns into account.

Response time measures at the page and survey levels have been shown to have limited utility compared to other measures of careless responding (Ashley & Shaughnessy, 2021; Ward & Meade, 2022); however, in line with previous research, the current study illustrated the usefulness of behavioral response data at the individual item level. Since careless responders are assumed to have rapid response times, which is consistent with their motivations to finish the survey quickly, the item response time is expected to catch careless responders regardless of their response style (Curran, 2016). This careless response detection method has potential utility for any self-report dataset with item-level response time.

In addition to the detection of careless responders for data quality purposes, the findings of the current study indicate that item response behavior provides clinically relevant information: emotional distress is correlated with a behavioral measure at an item-specific level. Researchers have drawn theoretical links between negative emotional states (i.e., depression, anxiety, and stress) and careless responding behavior (Ashley & Shaughnessy, 2021; Conjin et al., 2020). The cognitive and emotional characteristics of emotional distress are theorized as the mechanism explaining high rates of careless responding in emotionally distressed individuals (Ashley & Shaughnessy, 2021; Conjin et al., 2020). However, additional research is needed to clarify the links between careless responding and depression, anxiety, and stress. For example, some theories suggest that depressed individuals have heightened attention towards mood-congruent stimuli (e.g., survey items), while others suggest that avoidance of mood-congruent stimuli is typical in depressed individuals (Hubbard et al., 2016). The association between item response time and emotional distress found in the current study suggests that item response time can be a novel and precise approach to testing theories of emotional distress by unpacking the patterns of careless responding associated with specific emotional states (Castaneda et al., 2008, 2011; Snyder et al., 2015a, 2015b).

Finally, the current study found that machine learning trained on DASS-42 item response and response time can be another approach to predicting the severity of emotional distress, which corroborates previous research (e.g., Kumar et al., 2020; Priya et al., 2020; Srinath et al., 2022). While achieving 100% accuracy may not be feasible or necessary, the goal of using machine learning in this paper was not necessarily to outperform simple arithmetic summation, but rather to show the potential of a data-driven approach to analyze and predict emotional distress by considering both item response and response time. However, using machine learning based on item responses and response times can offer several advantages over the simple arithmetic summation of response data. First, it can allow for capturing subtle patterns that may not be evident through simple arithmetic summation. Second, machine learning techniques can handle complex interactions and dependencies among variables, which are very likely to be present in psychological assessment data such as DASS. Furthermore, these models have the potential to generalize to new datasets and populations, provided that they are trained on diverse and representative samples. This could enhance the applicability of the predictive models across different settings and populations, ultimately improving their utility in research contexts or other applied settings (e.g., healthcare).

Implications and Recommendations

Psychological measurement is in a period of advancement, with computerized testing affording researchers new ways to collect and interpret data. Item response time can be used to identify careless responding, and it has the potential to untangle the psychological mechanisms behind carelessness, particularly in those experiencing emotional distress. The relationship between careless responding and emotional distress, as well as the prediction of emotional states considering both item response and response time, can have important implications for cognitive psychology researchers. Thus, the current study has several research, psychometric, and applied implications and future directions for consideration.

First, researchers who use online self-report surveys can use the normative threshold method to flag careless responders, which allows the researchers to identify if these responders and their responses are influencing the data and obscuring important findings. This may be of particular importance among research pools in which the participants are receiving incentives to complete a survey. In such cases, participants may be motivated to finish the survey as quickly as possible to earn the incentive, leading to careless responses. This can result in low-quality data and invalid conclusions drawn from the survey results. If researchers are specifically using measures of emotional distress, such as the DASS-42, response time may help to identify those who have higher scores due to careless responding versus those who are genuinely emotionally distressed.

Second, the use of response time methods has clinical utility in assessing the profile validity of individuals who may be carelessly responding due to their emotional distress. Many commonly used psychological tests, including the DASS-42, do not include profile validity measures. One reason for this is that additional profile validity scales, such as positive impression management or defensiveness, add items to measures that are otherwise constructed to be as short and efficient as possible. Using response time to assess profile validity adds no additional items to these measures. Similar to profile validity measures, response time may also provide clinically relevant information to improve our understanding of the cognitive impairments that accompany emotional states. For example, item response time may help ascertain whether an emotionally distressed respondent tends to respond carelessly to avoid mood-congruent information and when they are biased to attend more carefully to mood-congruent information (Ashley & Shaughnessy, 2021; Hubbard et al., 2016). This is valuable information for treatment planning because useful interventions may vary based on whether a client over-attends to negative information (i.e., ruminates) or employs avoidance behavior. Thus, in terms of the first research question addressed in this study, understanding the relationship between careless responding and emotional distress can provide insight into the cognitive processes underlying emotional states.

Regarding the second research question, machine learning models can be utilized to analyze hidden patterns in both item response and response time for predicting self-reported measures of emotional states. Considering the current research that has demonstrated an association between emotional distress and response time, it is essential to incorporate response time into the assessment of emotional distress severity rather than relying solely on the arithmetic summation of responses. By doing so, we can achieve a better understanding of emotional distress.

Machine learning models that incorporate behavioral data, such as item response time, also have practical applications for adaptive testing. Wise (2020) suggested that the inclusion of item response data in adaptive testing allows test developers to provide a measure of attention to be considered in profile validity. With real-time monitoring of attention using item response data, developers can intervene to re-engage a respondent who is exhibiting careless responding. For example, if a respondent has several consecutive careless responses, a prompt may appear to remind them to carefully attend to each item. In educational contexts, response time has been included in adaptive testing models, but behavioral measures have not been widely used in computerized adaptive testing of personality and psychopathology.

Limitations and Future Research

There are several limitations in the current study. First, the use of large datasets is at higher risk of finding spurious correlations between variables. The current findings are situated within the theoretical and research literature supporting the assertion that cognitive symptoms of emotional distress impact response behavior, and thus provide greater confidence in the validity of the findings. Second, due to the nature of survey data, the direction of causation between careless responding and emotional distress cannot be confirmed. Other factors that were not studied in the current research (e.g., education, age, and formal diagnoses) may be confounding the relationship between emotional distress and careless responding. Similarly, the normative response method used to identify careless responses may be impacted by confounding variables, such as technical issues experienced by respondents, item wording, and item valence. Additional research is needed to understand how patterns of slow response time may be related to emotional distress due to low processing speed and poor concentration, and to differentiate these responses from slow, careful responders. The literature would benefit from a direct comparison of the normative threshold method with other detection methods (e.g., longstring).

There are also several limitations related to the second research question. First, there was a positive skew in our dataset with an overrepresentation of emotionally distressed responders. Kumar et al. (2020) noted the problem of determining the best predictive model when data is imbalanced between classification categories. Given the potential increase in computation time and considering the primary focus of our paper, we opted not to perform class balancing techniques. Future research can study different sampling methods, such as undersampling, oversampling, and ROSE techniques, to address the class imbalance. Second, the algorithm used in the machine learning model and its parameters may have impacted model performance. The current study used DT, RF, NB, and SVM algorithms, but future research can study other algorithms not included here. Third, machine learning models may make biased predictions for groups belonging to different demographic categories, such as race, gender, and age. Future research must consider these demographic variables to understand the generalizability of the model to different populations. Finally, to mitigate concerns of circularity, we split the data into separate training and validation sets, ensuring that predictive analyses were conducted on an independent dataset. Future research could further strengthen the validation of response-time-based indicators by examining their predictive utility using additional independent outcome measures.

The current study stimulates several future research directions. First, future research may specify which emotional states (e.g., depression, anxiety, excitement, and boredom) and demographics (e.g., age and gender) are associated with higher levels of carelessness for enhanced psychometric accuracy. This approach especially benefits individuals who score low on the DASS-42 due to poor insight into or masking of their emotional state but whose cognitive impairment is indicated by a behavioral measure, such as response time. Second, the identification of patterns in the data that are indicative of certain levels of emotional distress, which could help to improve the diagnostic accuracy of the DASS-42 scale, should be explored. For example, researchers can investigate whether a client's carelessness increases, ebbs and flows, or has a consistent rate throughout the survey. Thirdly, the normative response time approach may be applied within cognitive psychology and emotion regulation research to explore how different emotion regulation strategies impact cognitive processes and to compare response time cut scores amongst different populations. Finally, more research is needed to explore how machine learning models that include response time can be incorporated effectively into clinical and psychological assessments, such as adaptive testing and wellness-oriented smartphone applications.

Conclusion

Careless responding is a significant source of bias in online self-report surveys–a common data collection method in the field of psychology. The normative threshold method is an important approach for researchers to identify careless responders, and it circumvents the limitations of other approaches for detecting response bias. The normative threshold method also offers a novel behavioral measure for studying the impact of emotional distress on cognition. While all researchers using self-report measures hold responsibility and accountability for collecting valid and reliable data, those conducting research on emotional distress face additional validity threats because of the relationship between emotional distress and careless responding. The current study provides preliminary evidence that incorporating normative response thresholds into routine data cleaning practices and machine learning models may enhance the accuracy with which psychological researchers can describe and predict emotional distress.

References

American Psychological Association. (2020). *APA Guidelines for psychological assessment and evaluation.* APA Task Force on Psychological Assessment and Evaluation Guidelines. https://www.apa.org/about/policy/guidelines-psychological-assessment-evaluation.pdf

Antony, M. M., Bieling, P. J., Cox, B. J., Enns, M. W., & Swinson, R. P. (1998). Psychometric properties of the 42-item and 21-item versions of the Depression Anxiety Stress Scales in clinical groups and a community sample. *Psychological Assessment, 10*(2), 176–181. https://doi.org/10.1037/1040-3590.10.2.176

Arias, V. B., Garrido, L. E., Jenaro, C., Martínez-Molina, A., & Arias, B. (2020). A little garbage in, lots of garbage out: Assessing the impact of careless responding in personality survey data. *Behavior Research Methods, 52*(6), 2489–2505. https://doi.org/10.3758/s13428-020-01401-8

Ashley, M., & Shaughnessy, K. (2021). Predicting insufficient effort responding: The relation between negative thoughts, emotions, and online survey responses. *Canadian Journal of Behavioural Science / Revue canadienne des sciences du comportement, 55*(3), 198–209 https://doi.org/10.1037/cbs0000308

Beaudreau, S. A., & O'Hara, R. (2009). The association of anxiety and depressive symptoms with cognitive performance in community-dwelling older adults. *Psychology and Aging, 24*(2), 507–512. https://doi.org/10.1037/a0016035

Bowling, N. A., Huang, J. L., Bragg, C. B., Khazon, S., Liu, M., & Blackmore, C. E. (2016). Who cares and who is careless? Insufficient effort responding as a reflection of respondent personality. *Journal of Personality and Social Psychology, 111*(2), 218–

229. https://doi.org/10.1037/pspp0000085

Budiyanto, S., Sihombing, H. C., & Rahayu, I. M. F. (2019). Depression and anxiety detection through the closed-loop method using DASS-21. *Telkomnika, 17*(4), 2087–2097. https://doi.org/10.12928/TELKOMNIKA.v17i4.12619

Castaneda, A. E., Tuulio-Henriksson, A., Marttunen, M., Suvisaari, J., & Lönnqvist, J. (2008). A review on cognitive impairments in depressive and anxiety disorders with a focus on young adults. *Journal of Affective Disorders, 106*(1-2), 1–27. https://doi.org/10.1016/j.jad.2007.06.006

Castaneda, A. E., Suvisaari, J., Marttunen, M., Perälä, J., Saarni, S. I., Aalto-Setälä, T., Lönnqvist, J., & Tuulio-Henriksson, A. (2011). Cognitive functioning in a population-based sample of young adults with anxiety disorders. *European Psychiatry, 26*(6), 346–353. https://doi.org/10.1016/j.eurpsy.2009.11.006

Conijn, J. M., Emons, W. H. M., De Jong, K., & Sijtsma, K. (2015). Detecting and explaining aberrant responding to the Outcome Questionnaire-45. *Assessment, 22*(4), 513–524. https://doi.org/10.1177/1073191114560882

Conijn, J. M., Emons, W. H. M., Page, B. F., Sijtsma, K., Van der Does, W., Carlier, I. V. E., & Giltay, E. J. (2018). Response inconsistency of patient-reported symptoms as a predictor of discrepancy between patient and clinician-reported depression severity. *Assessment, 25*(7), 917–928. https://doi.org/10.1177/1073191116666949

Conijn, J. M., van der Ark, L. A., & Spinhoven, P. (2020). Satisficing in mental health care patients: The effect of cognitive symptoms on self-report data quality. *Assessment, 27*(1), 178–193. https://doi.org/10.1177/1073191117714557

Cuijpers, P., Li, J., Hofmann, S. G., & Andersson, G. (2010). Self-reported versus clinician-rated symptoms of depression as outcome measures in psychotherapy research on depression: A meta-analysis. *Clinical Psychology Review, 30*(6), 768–778. https://doi.org/10.1016/j.cpr.2010.06.001

Curran, P. G. (2016). Methods for the detection of carelessly invalid responses in survey data. *Journal of Experimental Social Psychology, 66*, 4–19. https://doi.org/10.1016/j.jesp.2015.07.006

Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: Attentional control theory. *Emotion, 7*(2), 336–353. https://doi.org/10.1037/1528-3542.7.2.336

Ferreri, F., Lapp, L. K., & Peretti, C. S. (2011). Current research on cognitive aspects of anxiety disorders. *Current Opinion in Psychiatry, 24*(1), 49–54. https://doi.org/10.1097/YCO.0b013e32833f5585

Forbey, J. D., Ben-Porath, Y. S., & Arbisi, P. A. (2012). The MMPI–2 computerized adaptive version (MMPI–2–CA) in a Veterans Administration medical outpatient facility. *Psychological Assessment, 24*(3), 628–639. https://doi.org/10.1037/a0026509

Funke, F. (2016). A web experiment showing negative effects of slider scales compared to visual analogue scales and radio button scales. *Social Science Computer Review, 34*(2), 244–254. https://doi.org/10.1177/0894439315575477

Gotlib, I. H., & Joormann, J. (2010). Cognition and depression: Current status and future directions. *Annual Review of Clinical Psychology, 6*, 285–312. https://doi.org/10.1146/annurev.clinpsy.121208.131305

Gross, J. J. (2015). Emotion regulation: Current status and future prospects. *Psychological Inquiry, 26*(1), 1–26. https://doi.org/10.1080/1047840X.2014.940781

Gummer, T., & Roßmann, J. (2015). Explaining interview duration in web surveys: A multilevel approach. *Social Science Computer Review, 33*(2), 217–234. https://doi.org/10.1177/0894439314533479

Guo, H., Rios, J. A., Haberman, S., Liu, O. L., Wang, J., & Paek, I. (2016). A new procedure for detection of students' rapid guessing responses using response time. *Applied Measurement in Education, 29*(3), 173–183. https://doi.org/10.1080/08957347.2016.1171766

Harms, C., Jackel, L., & Montag, C. (2017). Reliability and completion speed in online questionnaires under consideration of personality. *Personality and Individual Differences, 111*, 281–290. https://doi.org/10.1016/j.paid.2017.02.015

Heerwegh, D. (2003). Explaining response latencies and changing answers using client-side paradata from a web survey. *Social Science Computer Review, 21*(3), 360–373. https://doi.org/10.1177/0894439303253985

Hintze, J. M., & Silberglitt, B. (2005). A longitudinal examination of the diagnostic accuracy and predictive validity of R-CBM and high-stakes testing. *School Psychology Review, 34*, 372–386. http://doi.org/10.1080/02796015.2005.12086292

Huang, J. L., Curran, P. G., Keeney, J., Poposki, E. M., & DeShon, R. P. (2012). Detecting and deterring insufficient effort responding to surveys. *Journal of Business and Psychology, 27*(1), 99–114. https://doi.org/10.1007/s10869-011-9231-8

Hubbard, N. A., Hutchison, J. L., Turner, M., Montroy, J., Bowles, R. P., & Rypma, B. (2016). Depressive thoughts limit working memory capacity in dysphoria. *Cognition and Emotion, 30*, 193–209. https://doi.org/10.1080/02699931.2014.991694

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2017). *An introduction to statistical learning with applications in R*. New York: Springer.

Jones, A., Earnest, J., Adam, M., Clarke, R., Yates, J., & Pennington, C. R. (2022). Careless responding in crowdsourced alcohol research: A systematic review and meta-analysis of practices and prevalence. *Experimental and Clinical Psychopharmacology, 30*(4), 381–399. https://doi.org/10.1037/pha0000546

Keeley, J. W., Webb, C., Peterson, D., Roussin, L., & Flanagan, E. H. (2016). Development of a response inconsistency scale for the personality inventory for DSM–5. *Journal of Personality Assessment, 98*(4), 351–359. https://doi.org/10.1080/00223891.2016.1158719

Kilgus, S. P., Chafouleas, S. M., & Riley-Tillman, T. C. (2013). Development and initial validation of the Social and Academic Behavior Risk Screener for elementary grades. *School Psychology Quarterly, 28*(3), 210–226. https://doi.org/10.1037/spq0000024

Kong, X. J., Wise, S. L., & Bhola, D. S. (2007). Setting the response time threshold parameter to differentiate solution behavior from rapid-guessing behavior. *Educational and Psychological Measurement, 67*(4), 606–619. https://doi.org/10.1177/0013164406294779

Kumar, P., Garg, S., & Garg, A. (2020). Assessment of anxiety, depression and stress using machine learning models. *Procedia Computer Science, 171*, 1989–1998. https://doi.org/10.1016/j.procs.2020.04.213

LePagea, J. P., Mogge, N. L., & Sharpe, W. R. (2001). Validity rates of the MMPI-2 and PAI in a rural inpatient psychiatric facility. *Assessment, 8*(1), 67–74. https://doi.org/10.1177/107319110100800106

Liu, Y. S., Song, Y., Lee, N. A., Bennett, D. M., Button, K. S., Greenshaw, A., Cao, B., & Sui, J. (2022). Depression screening using a non-verbal self-association task: A machine-learning based pilot study. *Journal of Affective Disorders, 310*, 87–95. https://doi.org/10.1016/j.jad.2022.04.122

Lovibond, S.H., & Lovibond, P.F. (1995). *Manual for the Depression Anxiety Stress Scales* (2nd ed.). Sydney: Psychology Foundation.

McNair, D. M., Lorr, M., & Droppleman, L. F. (1971). *Manual for the Profile of Mood States*. San Diego, CA: Educational and Industrial Testing Services.

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods, 17*(3), 437–455. https://doi.org/10.1037/a0028085

Open Source Psychometrics Project. (2019). *Open psychology data: Raw data from online personality tests.* https://openpsychometrics.org/_rawdata/

Priya, A., Garg, S., & Tigga, N. P. (2020). Predicting anxiety, depression and stress in modern life using machine learning algorithms. *Procedia Computer Science, 167*, 1258–1267. https://doi.org/10.1016/j.procs.2020.03.442

Rios, J. A., & Soland, J. (2021). Parameter estimation accuracy of the effort-moderated item response theory model under multiple assumption violations. *Educational and Psychological Measurement, 81*(3), 569–594. https://doi.org/10.1177/0013164420949896

Salthouse T. A. (2012). How general are the effects of trait anxiety and depressive symptoms on cognitive functioning? *Emotion, 12*(5), 1075–1084. https://doi.org/10.1037/a0025615

Schnipke, D. L., & Scrams, D. J. (1997). Modeling item response times with a two-state mixture model: A new method of measuring speededness. *Journal of Educational Measurement, 34*(3), 213–232. https://doi.org/10.1111/j.1745-3984.1997.tb00516.x

Snyder, H. R., Kaiser, R. H., Warren, S. L., & Heller, W. (2015). Obsessive-compulsive disorder is associated with broad impairments in executive function: A meta-analysis. *Clinical Psychological Science, 3*(2), 301–330. https://doi.org/10.1177/2167702614534210

Snyder, H. R., Miyake, A., & Hankin, B. L. (2015). Advancing understanding of executive function impairments and psychopathology: Bridging the gap between clinical and cognitive approaches. *Frontiers in Psychology, 6*, 328. https://doi.org/10.3389/fpsyg.2015.00328

Srinath, K. S., Kiran, K., Pranavi, S., Amrutha, M., Shenoy, P. D., & Venugopal, K. R. (2022). *Prediction of depression, anxiety and stress levels using Dass-42* [Paper presentation]. 2022 IEEE 7th International Conference for Convergence in Technology (I2CT), Mumbai, India, 1–6.

Sun, M. K., & Alkon, D. L. (2014). Stress: Perspectives on its impact on cognition and pharmacological treatment. *Behavioural Pharmacology, 25*(5–6), 410–424. https://doi.org/10.1097/FBP.0000000000000045

Tada, M., Uchida, H., Suzuki, T., Abe, T., Pollock, B. G., & Mimura, M. (2014). Baseline difference between patients' and clinicians' rated illness severity scores and subsequent outcomes in major depressive disorder: Analysis of the sequenced treatment alternatives to relieve depression data. *Journal of Clinical Psychopharmacology, 34*(3), 297–302. https://doi.org/10.1097/JCP.0000000000000112

van der Linden, W. J., Klein Entink, R. H., & Fox, J. P. (2010). IRT parameter estimation with response times as collateral information. *Applied Psychological Measurement, 34*(5), 327–347. https://doi.org/10.1177/0146621609349800

Ward, M. K., & Meade, A. W. (2023). Dealing with Careless Responding in Survey Data: Prevention, Identification, and Recommended Best Practices. *Annual Review of Psychology, 74*, 577–596. https://doi.org/10.1146/annurev-psych-040422-045007

Wardenaar, K. J., Wanders, R. B. K., Roest, A. M., Meijer, R. R., & de Jonge, P. (2015). What does the Beck Depression Inventory measure in myocardial infarction patients? A psychometric approach using item response theory and person-fit. *International Journal of Methods in Psychiatric Research, 24*(2), 130–142. https://doi.org/10.1002/mpr.1467

Watson, D., & Clark, L. A. (1994). The PANAS-X: *Manual for the positive and negative affect schedule - expanded form.* University of Iowa.

Wise, S. L. (2006). An investigation of the differential effort received by items on a low-stakes, computer-based test. *Applied Measurement in Education, 19*(2), 25–114. https://doi.org/10.1207/s15324818ame1902_2

Wise, S. L. (2020). An intelligent CAT that can deal with disengaged test taking. In H. Jiao & R. W. Lissitz (Eds.), *Application of Artificial Intelligence to Assessment* (pp. 161–174). Information Age Publishing, Inc.

Wise, S. L., & Kong, X. (2005). Response time effort: A new measure of examinee motivation in computer-based tests. *Applied Measurement in Education, 18*(2), 163–183. https://doi.org/10.1207/s15324818ame1802_2

Wise, S. L., & Ma, L. (2012, April). *Setting response time thresholds for a CAT item pool: The normative threshold method* [Paper presentation]. Annual Meeting of the National Council on Measurement in Education, Vancouver, British Columbia, Canada.

Zuckerman, M., & Lubin, B. (1985). *Manual for the multiple affect adjective check list.* San Diego: Educational and Industrial Testing Service.

Figure 1

Methodological Framework for Addressing Research Question 1

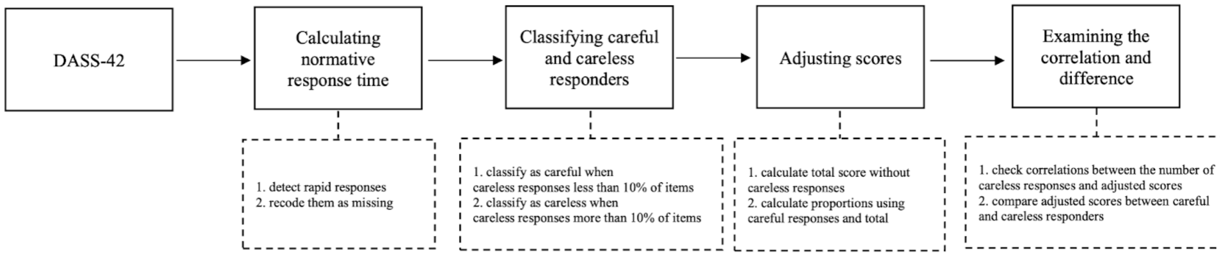


Figure 2

Methodological Framework for Addressing Research Question 2

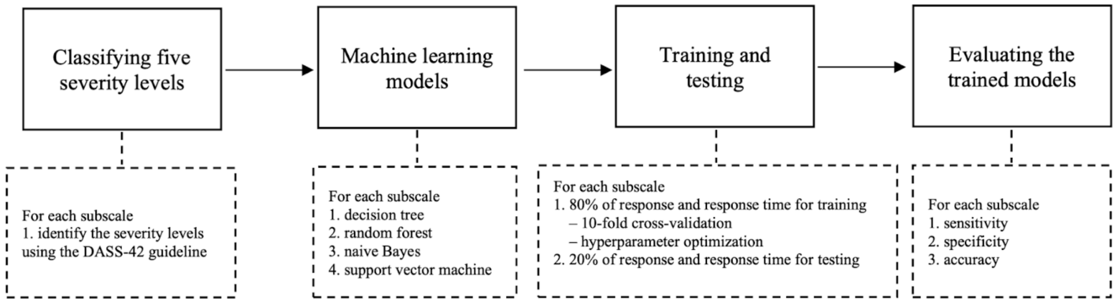


Figure 3

Comparison of DASS-42 Subscale Scores between Careful and Careless Responders

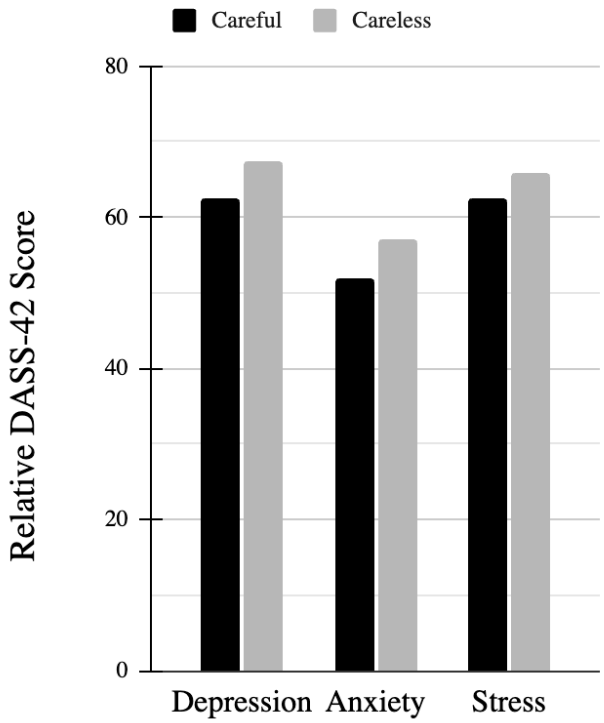
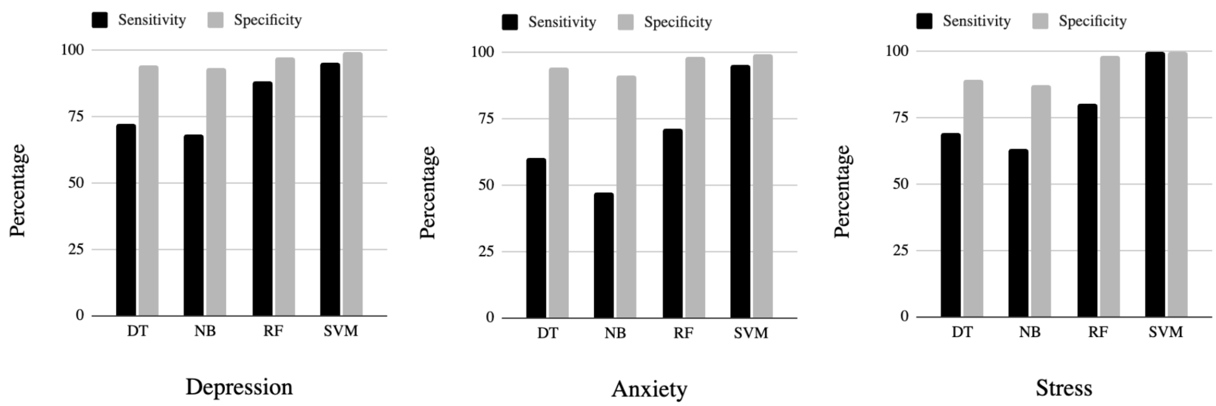


Figure 4

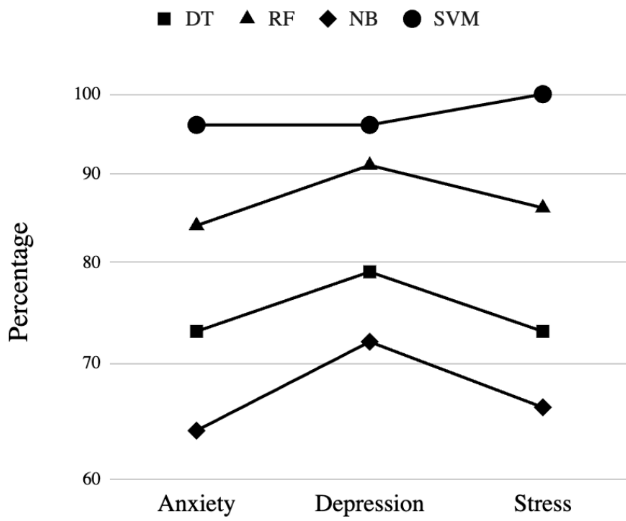
Macro-averaged Classification Metrics for Emotional Distress



Note. DT: decision tree; RF: random forest; NB: naive Bayes; SVM: support vector machine.

Figure 5

Accuracy of Classification for Machine Learning Models



Note. DT: decision tree; RF: random forest; NB: naive Bayes; SVM: support vector machine

Table 1

Sociodemographic Summary

Demographic variable	*n*	%
Education		
Less than high school	3881	10
High school	14325	38
University degree	14399	38
Graduate degree	4729	13
Urbanicity		
Rural	7892	21
Suburban	12595	33
Urban	16972	45
Gender		
Male	8365	22
Female	28864	76
Other	528	1
Ethnicity		
Asian	21910	58
Arab	311	1
Black	575	1
Indigenous Australian	23	<1
Native American	209	1
White	10236	27
Other	4562	12
Marital status		
Never married	32472	86
Currently married	4122	11
Previously married	1039	3

Table 2

Guide for Severity Levels of Emotional Distress in DASS-42

	Depression	Anxiety	Stress
Normal	0–9	0–7	0–14
Mild	10–13	8–9	15–18
Moderate	14–20	10–14	19–25
Severe	21–27	15–19	26–33
Extremely severe	28+	20+	34+

Table 3

Single Classification Metrics for Emotional Distress by Each Level

	Normal	Mild	Moderate	Severe	Extremely severe
			Depression		
DT					
Sensitivity	91.20	48.03	68.14	64.38	92.17
Specificity	96.67	95.34	93.48	93.40	95.06
RF					
Sensitivity	97.50	59.97	90.48	86.39	97.55
Specificity	97.93	98.75	96.03	97.71	**98.28**
NB					
Sensitivity	79.85	57.30	52.32	70.42	81.78
Specificity	98.28	91.69	91.33	87.64	97.50
SVM					
Sensitivity	**98.22**	**87.64**	**94.38**	**92.50**	**98.85**
Specificity	**99.37**	**99.56**	**98.68**	**99.21**	97.66

Table 3 (continued)

			Anxiety		
DT					
Sensitivity	88.55	22.84	58.27	45.64	88.42
Specificity	95.24	96.27	89.83	90.82	92.70
RF					
Sensitivity	97.73	4.80	86.62	64.28	96.60
Specificity	95.96	**99.77**	91.53	96.34	95.62
NB					
Sensitivity	92.11	0.96	52.03	12.43	83.38
Specificity	85.31	99.73	82.56	95.60	88.95
SVM					
Sensitivity	**98.22**	**87.64**	**94.38**	**92.50**	**98.85**
Specificity	**99.37**	99.56	**98.68**	**99.21**	**97.66**

Table 3 (continued)

	Normal	Mild	Moderate	Severe	Extremely severe
			Stress		
DT					
Sensitivity	88.70	41.79	65.86	69.47	80.51
Specificity	94.72	93.19	89.21	91.36	97.06
RF					
Sensitivity	96.20	48.70	89.49	90.90	87.31
Specificity	96.39	97.95	93.06	95.68	99.55
NB					
Sensitivity	79.10	45.25	54.92	62.03	78.33
Specificity	96.17	89.41	87.11	89.43	95.54
SVM					
Sensitivity	**99.96**	**100.0**	**99.82**	**100.0**	**99.91**
Specificity	**100.0**	**99.97**	**99.98**	**99.97**	**100.0**

Note. DT: decision tree; RF: random forest; NB: naive Bayes; SVM: support vector machine. Bold indicates the highest, with 70% acceptable and 80% optimal thresholds in each subscale.

Pretty Privilege vs. Ingroup Bias in Decision Making

Vicky Zhu & Daniel White
University of Technology Sydney, Psychology, Faculty of Health, Sydney, NSW

In-group and attractiveness bias are well-established concepts in social psychology. This study examines the concurrent influence of these concepts on the decision-making process using the Minimal Group Paradigm. Confederates, individuals who appear to be participants, were used to simulate out-group members. Participants (n = 119, aged 20–30 years) answered a series of mathematics questions, followed by a response, agreeing or disagreeing with the participant's answer, from a confederate. Participants were then asked to rate the attractiveness of the confederates. Results indicated that in-group bias significantly outweighed attractiveness bias. Participants changed answers more frequently when their group disagreed, regardless of the confederate's attractiveness. Results highlighted the robust effects of group membership on decision-making. Additional research is required to explore confounds within decision-making, such as individual differences and familiarity bias.
Keywords: in-group bias, attractiveness bias, decision-making, minimal group paradigm, group dynamics

Throughout our lives, we face situations that require us to make appropriate and sound decisions. Often, people make decisions under the misconception that their choices are purely the result of objective and conscious processing. Nonetheless, research indicates that we are susceptible to external influences, including attractiveness and in-group bias (Mackie & Ahn, 1998; Voit et al., 2021). Attractiveness and in-group bias may lead individuals to form inaccurate beliefs, which in turn influence decision-making. Attractiveness bias refers to the tendency for individuals to positively view attractive people, solely on the basis of their physical appearance (Shahani et al., 1993). Meanwhile, in-group bias refers to individuals' tendency to view members of their own group positively, which is driven by perceived group membership (Knoblock-Westerwick et al., 2020). Studies examining attractiveness bias and its influence in a group setting have shown that attractive individuals are stereotyped as being more trustworthy, leading to favoritism within groups (Cellerino, 2013). Although research has considered the impact of attractiveness within in-groups, little research has evaluated their simultaneous impact. When individuals face a conflict between appealing to attraction and in-group bias, it is unknown which factor has a greater influence on the decision-making process. The current study aims to explore the outcome of these conflicting influences on individuals' decisions.

In-Group Bias

In-group bias describes the tendency of an individual to favor members of their own group (Scheepers et al., 2006). For early humans, the development of collaborative social systems (i.e., division of labor and trade) ensured survival (Baumeister et al., 2015). Although salient, group membership results in out-group members being less persuasive as compared to in-group members (Mcgarty et al., 1994). When examining the influence of national in-group identity on persuasiveness, Adam-Troian et al. (2020) found that when national group identities were made salient, negative attitudes towards ethnic minorities (out-group) were more prevalent. Furthermore, it has been established that humans are innately drawn to their own group and display out-group xenophobia (Tidwell et al., 2017). These traits have historically led to greater survival rates.

In-group bias has been prevalent since ancestors formed closely connected groups that competed with others for survival (Hare, 2017). Presently, group adhesion is central across various fields, including education and conventional workplaces (Rathbone et al., 2023). When examining how individuals behave in the presence of other in-group members, Castelli et al. (2008) found that individuals exhibited more egalitarian responses when their behaviors aligned with the group's social norms. Additionally, while egalitarian group members were not subject to active discrimination, in-group members who exhibited behaviors aligned with the group's collective interest received more favorable evaluations (Castelli et al., 2008). If an individual expressed negative attitudes toward egalitarian members, it could diminish the group's overall cohesion and potential contributions. Consequently, individuals are less inclined to engage in prejudiced behavior and are more likely to make decisions they perceive as beneficial to the group's welfare (Terry et al., 2000). This is supported by Terry et al. (2000), which looked at attitudes and behaviors in terms of in-group bias and decision-making. Participants took part in a mock-jury paradigm in which they were presented with a hypothetical case and asked to render a verdict based on the evidence provided. Participants were more likely to make decisions consistent with their attitudes when they aligned with the in-group's attitude (Terry et al., 2000).

Many organisations thrive on teamwork, as performance is largely dependent on group members' ability to work together towards a common goal (Espín et al., 2019). Consequently, groups may be hindered or facilitated by in-group bias, underscoring the importance of identifying when it is present. Chai et al. (2022) showed that in-group bias was present between the ages of 5 and 6. When given a fictional story about an in-group/out-group member's sharing behavior, researchers predicted that in-group members would share more than their out-group peers. As such, it is worth noting that these biases became stronger in different contexts, and achieving intergroup collaboration can depend on how well in-group biases are managed (Li et al., 2021). When Li et al. (2021) investigated fate control — the belief that the future is largely predetermined but can be influenced by individuals' actions — on in-group bias during the COVID-19 pandemic, it was found that higher fate control was associated with higher risk perception, and this led to stronger in-group bias in donating to help with COVID-19. However, such behaviors change when resources are scarce. Cui et al. (2023) looked at how participants allocated resources to themselves and an in-group/out-group. The in-group bias was more prevalent in the scarcity condition, where participants were observed to allocate more resources to in-group members. These studies highlight situations when in-group bias increases, allowing either to reduce its harm or use it for positive outcomes.

In-group bias can manifest in the tendency to agree with the group's opinions (Scheepers et al., 2006). This can be seen through members favoring members of their own group, with individuals being more likely to select attractive people as being supporters of their own party (Nicholson et al., 2016). In addition, Knippenberg and Wilke (1992) investigated the effects of framing arguments on persuasiveness. Participants were presented with arguments that either aligned or contradicted their group's attitudes and arguments. Results show that individuals were most likely to agree with the argument that best aligned with their group's views. Another expression of in-group bias is evident in the tendency to agree with group decisions depending on the salience of group identity. Skinner and Stephenson (1981) explored how highlighting group identity and contrasting in-group views with out-group views influences individuals' expression of their opinions. It was found that emphasising group affiliation led to participants intensely expressing their opinions, agreeing with group views. These findings highlight how in-group bias shapes individual decision-making.

Formation of a "Group"

It is widely known that individuals are more likely to cooperate with others with whom they share a group identity (Calanchini et al., 2022). However, recent evidence suggests that prior interaction may not be necessary to identify them as an in-group member (Kurzban et al., 2001). Instead, humans use symbolic cues and other ways to elicit favoritism towards in-group members. Hence, to create a group in laboratory settings, it is necessary to understand the Minimal Group Paradigm, which consists of three main components (Otten, 2016). Firstly, categorization must be novel and arbitrary, meaning there must be no history of experiences with any in-group or out-group. Second, categorization is anonymous; participants are to have no face-to-face interaction with group members. Lastly, there are no direct relationships between allocations and self-interest. A minimal group designed correctly should evoke behaviors where individuals favor their own group members.

The Minimal Group Paradigm

Montalan et al. (2012) examined empathy towards in-group/out-group members using the Minimal Group Paradigm. Participants were split into two groups using a dot-estimation task: the underestimators (who allegedly underestimated the number of dots shown) and the overestimators (who allegedly overestimated the number of dots shown). Due to the fictive nature of the groups, participants would have had no previous experiences with the in-group or out-group. After participants were split into groups, participants did a pain stimulator task individually, having no face-to-face interactions with other members. Results showed evidence of in-group bias: participants made decisions to show more empathy for those in their own group in comparison to those in the out-group.

In another study, Reynolds et al. (2007) looked at predisposing factors that could lead to discrimination. Participants were assigned to minimal groups. Participants in the random condition were told that group allocations were done on a completely random basis. On the other hand, participants in the voluntary condition were asked to circle the group they wanted to be in. After the group assignment, participants completed various personality measures. Results showed behaviors expected of an in-group member, with in-group identification alone being the strongest predictor of participants' tendency to discriminate.

Attractiveness

Attractiveness plays an undeniable role in evolutionary success, with attractive traits being associated with potentially fitter offspring (Gangestad & Scheyd, 2005). Although it is difficult to standardize an attractive face, certain features are preferred. Attractive women often had fuller lips, high-arched eyebrows, smaller and more tapered noses, and less angular jaws (Pflüger et al., 2012). These feminine features were associated with disease resistance, high estrogen levels, and fertility (Muñoz-Reyes et al., 2014). In men, those with large eyes, prominent cheekbones, large chins, big smiles, and high-status clothing were considered attractive (Cunningham et al., 1990).

Attractiveness Bias

Attractiveness plays a role in social encounters, as many first impressions are based on visual features (Nordholm, 1980). Attractive people are often associated with positive personality traits, such as being sociable and intelligent (Tsukiura & Cabeza, 2010). Whereas non-attractive individuals are characterised by negative personality traits, such as being less altruistic and less intelligent. In their study of attractiveness and selective bias in attributing moral character, Tsukiura and Cabeza (2010) found that attractiveness biases hinder perceptions of others. Participants were more likely to attribute moral states to attractive rather than less attractive individuals. In addition, attractive individuals are more highly rated in perceived warmth in comparison to non-attractive individuals (Klebl et al., 2021). This is due to the beauty-is-good (BIG) stereotype.

The BIG stereotype can be explained using the halo effect: the positive evaluation of one trait influences the evaluation of other unrelated characteristics (Klebl et al., 2021). Individuals perceive attractive people as more correct, and this can lead to the desire to conform to their decisions. Batres and Shiramizu (2022) looked at attractiveness having a "halo effect," where people associated social desirability with attractive individuals and provided evidence of the halo effect cross-culturally. Hence, simply being attractive produces a "halo" effect that increases the chances of other people assigning positive traits to them, impacting the decision-making process.

Being perceived as attractive offers advantages in various aspects of life. In relationships, many attractive people report being more satisfied with their dating life (Berscheid et al., 1971). In a job setting, it was found that non-attractive job candidates need to submit 33% more applications in comparison to their attractive counterparts (Maurer-Fazio & Lei, 2014). Additionally, unattractive individuals were found to earn 7–10% less than average-looking individuals (Hamermesh & Biddle, 1993). As well as this, attractive people were found to pay lower bail and fine amounts for misdemeanor charges (Downs & Lyons, 1991). In Griffin and Langlois' (2006) study, the advantages of attractiveness and the disadvantages of unattractiveness were examined. The results suggested that being unattractive has disadvantages, with adults and children both giving low scores of positive attributes (sociability, altruism, and intelligence) to unattractive individuals. Another study also found that more physically attractive individuals were viewed in both a positive light and more accurately in first impressions (Lorenzo et al., 2010). Individuals forming quick impressions often lead to inaccurate judgments, and these biases influence the decision-making process in many different social situations (Zuckerman et al., 1995).

People often make inferences about individuals' character by judging their facial appearance (Øvervoll et al., 2020; Willis & Todorov, 2006). As exposure to individuals with attractive faces increases due to media portrayal, it is important to know how often this occurs. In Herbozo et al.'s (2004) study, where three raters coded messages present in children's videos, it was found that 72% of the analyzed videos emphasized physical attractiveness. Additionally, 84% of the videos associated female attractiveness with sociability and kindness. The results from this study indicate the prevalence of attractiveness stereotypes, hence the need to understand how this can affect decision-making. Willis and Todorov (2006) investigated the conditions in which participants are likely to make inferences based on facial appearance. It was found that merely exposing participants to images for 100 milliseconds was enough for judgments of attractiveness to be made.

Additionally, longer exposure time was found not to affect participants' judgement of characteristics. Such results show that exposing participants to just the face region is sufficient in generating attraction judgements.

The Chicago Face Database (CFD) was created using face morphing software and provides sets of faces and norming data (Ma et al., 2015). The benefits of using the CFD include access to a large variety of faces (Ma et al., 2020). The availability of large databases of faces offers the potential for valuable contributions to the field of psychology. The CFD has previously been used in various research done on appearances (Freud et al., 2020; Landy et al., 2020; Marini et al., 2021).

Impact of Attractiveness on In-Group Bias

Despite the concrete evidence on in-group bias, recent research suggests that group membership biases are malleable (Rudman et al., 2001). Rudman et al.'s (2001) study demonstrated that affective and cognitive processing play a role in reducing prejudice. Furthermore, individuals' prejudices are dependent on their biases as well as the situation. Despite compelling evidence of the impact of in-group bias, the influence is not absolute. Dang et al.'s (2019) study was conducted to look at the impact of criticism on in-group favoritism. When participants were criticized by an authoritative outside figure, it was found that in-group favoritism decreased. It was proposed that these results were due to the authoritative figures being viewed with admiration and possessing qualities that are considered attractive. However, the study did not investigate its impact on the subsequent decision-making of the group members. Additionally, those who were experiencing a threat to their social self-esteem (their in-group) allocated fewer resources to their own group members, showing a decrease in behavior consistent with in-group bias. However, substantial research is still required on the interaction between attractiveness and group membership. In another study, Kniffin et al. (2014) explored how perceptions of leader attractiveness are influenced by group membership. The results from this study were that in-group leaders were rated as more attractive than out-group leaders. This finding suggests that, despite similar levels of familiarity with both leaders, the out-group leader's attractiveness was not stronger than the in-group bias.

The impact of attractiveness on groups has been recognized as influencing people's feelings and behaviors (Krendl et al., 2011). A study looked at sorority recruitment and what factors influenced acceptance of potential members into high-status and low-status groups. The results revealed that high-status sororities prioritized the attractiveness of participants during the decision-making process. These results highlight the complexity behind the effect of attractiveness on social groups, and the increased likelihood of accepting a member depending on attractiveness.

Individuals affiliate with others who are like themselves, and many groups are formed by a shared interest in certain topics (Chen & Kenrick, 2002). A study was conducted to investigate the effects of group membership on the perception of others. Results showed that individuals assumed that other members of the in-group share attitudes similar to their own, and that out-group members tend to have attitudes dissimilar to their own. Interestingly, it was found that when a member of the out-group shares similar attitudes, participants become more attracted to them, especially when the out-group has negative stereotypes. These results suggest that although an individual may be part of the out-group, it is possible that there are other external factors that can change how an in-group member perceives them and consequential decision-making. Given that most research looks at how in-group bias interacts with attractiveness, there is a need to examine scenarios wherein they contradict each other.

While there have been many independent studies on attractive and in-group bias, further research is needed to investigate their conflicting influence. Results studying these two factors affect our understanding of decision-making and the need to consider attractiveness while forming decision-making groups (e.g., jury panels, debating teams, and soccer teams). With research suggesting that in-group bias is not as resilient as previously believed, there is a possibility that the attractiveness of an out-group member will affect how strong in-group behaviors are. Furthermore, results further support the use of the CFD in future experiments. Further research is required to delve into decision-making and how attractiveness can interfere with in-group bias to affect this process. Hence, the purpose of the current experiment is to investigate whether attractiveness influences in-group bias. It is hypothesised that the presence of an attractive out-group member will overpower in-group bias and participants will be more likely to change their answer when they deem the out-group member attractive, regardless of their group's answer.

Method

Participants

The convenience sample consisted of 119 University of Technology Sydney (UTS) students aged 20 to 30. Recruitment was conducted through online postings on the SONA system, and participants were reimbursed with 0.5 credits for their participation.

Ethical Clearance

Ethical clearance to conduct research with human subjects was obtained through UTS Psychology Low and Negligible Risk (LNR) Ethics Panel (G-15-2023; see Appendix A). Informed consent from participants was obtained through a Participation Information Sheet provided prior to the commencement of the study. Participants were informed that they could withdraw consent at any time without penalty.

The CFD

The study utilized the CFD (Ma et al., 2015). For the purposes of the current experiment, only eight faces were chosen as the confederates. Using the norming data provided, the chosen faces had to be between the ages of 20 and 30, as this reflected the ages of the participants. This allows for controlling age as a confounding variable for perceived attractiveness, as it has previously been shown that younger people are considered more attractive (Zebrowitz & Franklin, 2014). For the CFD, attractiveness was measured on a 7-point Likert scale (1 = *least attractive* to 7 = *most attractive*). The images used (confederates) were chosen from the extreme ends of the given attractiveness ratings. Four of the most highly rated faces were chosen as the attractive confederates. This included faces with the ratings of 5.48, 5.31, 5.24, and 5.12. In addition, four of the lowest-rated faces were chosen as the non-attractive confederates. This included faces with the ratings of 1.61, 1.55, 1.54, and 1.52. Attractiveness ratings and ages were the only two factors considered when choosing confederates.

The Minimal Group Paradigm

To stimulate in-group bias in a short period of time, the current study utilized the Minimal Group Paradigm (Otten, 2016). Creating a new group was necessary as it eliminates the confounding effect of biases on results. This means that any preconceptions and previous experiences would not influence participants' responses.

It has been established that prior interaction may not be necessary for participants to identify themselves as an in-group member (Kurzban et al., 2001). Additionally, it has been found that merely categorizing individuals into two social groups is enough to elicit behaviors of group members.

A novel or arbitrary categorization means participants had no prior experiences with the in-group or out-group. The current experiment addresses this through participants being randomly assigned to groups, as well as them being informed that their groups are based on their study sign-up time. This also ensured that participants had no relationship between their allocated group and their self-interests. Anonymous categorization means participants do not have any face-to-face interactions with in-group or out-group members. The online nature of the experiment addresses this. Participants' identities were kept anonymous, and they did not encounter other participants. A minimal group designed correctly will evoke behaviors of in-group favoritism (Hertel & Kerr, 2001).

This study utilized the Minimal Group Paradigm in an online format. Janneck et al. (2013) conducted an experiment to replicate the Minimal Group Paradigm in an online setting. Arbitrary groups were created, and varying degrees of information about other group members were made available to participants. In both informal and work settings, participants showed in-group favoritism. Additionally, when less information was made available to participants, in-group bias was more prominent. This study exhibits the effectiveness of Minimal Group Paradigms in online settings.

Online Groups

The current study was conducted online via Qualtrics. This was due to convenience, time constraints, financial constraints, and being a standard practice for research conducted with UTS. Furthermore, online environments allow for the convenience in creating and conducting experiments involving manipulation online if the task does not require any physical presence (Horton et al., 2010). This environment ensured participant anonymity while allowing for controlled deception regarding the existence of other peers.

Group Allocation

Participants were informed that there were groups as well as independent individuals. However, the participants were unaware that they had been placed in

Group A and completed the study individually at different times. They were informed that their group allocation was dependent on the time frame in which they signed up for the study. As per the Minimal Group Paradigm, categorization was novel, and there was no relationship between allocations and participants' self-interests.

Testing Phase

To ensure that participants answered honestly, they were informed that their responses and choices would be anonymous (Schitter et al., 2019). Participants were asked to complete eight multiple-choice mathematics questions (see Appendix A). They were given 10 seconds to answer each question. None of the questions had correct answers. The intention was to create uncertainty within the participants regarding the correct answer.

After the participants answered each question, they were shown a summary of their group's (Group A) response and a response from a confederate (see Appendix B). The responses from the group and the confederate were predetermined and not dependent on active participation. The eight possible responses were randomized (see Table D1). Participants were then given 10 seconds to decide if they wanted to change or retain their answer. This was repeated eight times.

Rating Phase

After participants had completed the mathematics questions and made their decision to change or maintain their original question, they were asked to state whether they think the person in the image is attractive in a yes/no questionnaire (see Appendix C). Although the more "attractive" faces from the CFD have been included, it is important to factor in individual preference. Participants were again reminded that their responses were kept anonymous.

Calculator Use and Debrief

Participants with incomplete responses and participants who responded "Yes" when asked whether they used a calculator had their responses removed. A total of 32 responses were removed. Participants were then informed of the manipulation being used. They were debriefed as per UTS requirements.

Results

Data Management

Data was collected on Qualtrics, and a copy of the data was placed on UTS's OneDrive as per UTS's requirements. Data was de-identified and password-protected. Data was de-individualized and exported onto a personal laptop to conduct analysis, and was deleted promptly after data analysis was complete.

Statistical Test and Hypothesis

The relationship between in-group bias and attractiveness bias was examined using a Chi-squared test of independence. This was done by looking at participants' likelihood of changing their answer when presented with the group and an attractive/unattractive individual who agreed or disagreed with their answer.

Chi-Square Test of Independence

Attractiveness of Confederate vs. Answer Change

The results of the Chi-squared test of independence revealed no significant difference between answer change and attractiveness, $X^2(1, N = 696) = 2.12$, $p = .145$. There was some variation between the observed and expected scores (Table D2 in Appendix D), and participants were more likely to retain their answer when the confederate was unattractive (see Figure 1). However, the Chi-squared test of independence indicated that these results were non-significant ($p > .05$). Cramér's V of .06 revealed little, if any, association between the two variables.

Group Response vs. Answer Change

Results from the Chi-squared test of independence revealed a significant relationship between group response and answer change, $X2(1, n = 696) = 20.80$, $p < .001$. There was variation between observed and expected results (Table D3 in Appendix D). Participants were more likely to change their answer when the group disagreed with their answer (see Figure 2). Cramér's V value of .17 revealed the strong relationship between the two variables.

Attractiveness vs. Answer Change When Group Agrees and Confederate Disagrees

The Chi-squared test of independence was used to examine the relationship between answer change and attractiveness bias when the group agreed, and the confederate disagreed with the participants' answer. It showed that the relationship between these two variables was non-significant, $X^2(1, n = 348) = 1.88$, $p = 1.70$. Although there was some variation in observed (Figure 3) and expected values (Table D4 in Appendix D), Cramér's V of 0.07 revealed the low association between answer change and attractiveness of the confederate when the group agrees, and the confederate disagrees with participants' answers.

The relationship between the answer change and attractiveness when the group disagreed and the confederate agreed with the participant's answer showed no significance, $X^2(1, n = 348) = 0.34, p = .562$. There were variations in observed and expected scores (Table D5 in Appendix D), and the likelihood of retaining an answer was observed (Figure 4), but the Chi-square test of independence showed that there was no significant difference between conditions. The Cramér's V of .03 revealed little to no significance between answer change and attractiveness in this context.

Discussion

The aim of the study was to examine whether attractiveness influences in-group bias. More specifically, the present study investigated whether the presence of an attractive out-group member would override the effects of in-group bias in decision-making. Contrary to the initial prediction that participants would be more likely to change their answer in the presence of an attractive out-group member, the results of the study revealed no significant differences. This suggests that attractiveness does not affect in-group bias. However, group responses had a significant effect on whether participants changed their answers. More specifically, when the group disagreed with the participant's answer, participants were more likely to change their answer, irrespective of the attractiveness of the out-group member (confederate). This suggests that in-group bias is stronger than the attractiveness bias. Such results highlight that the Minimal Group Paradigm is a valid approach to developing a sense of cohesion within individuals.

There was a lack of significant findings when looking at:

1. Individuals' ratings of attractiveness versus answer change.
2. Individuals' ratings of attractiveness versus answer change when the group agrees and the confederate disagrees.
3. Individuals' ratings of attractiveness versus answer change when the group disagrees and the confederate agrees.

The current study explored the contradicting effects of attractiveness and in-group bias, and the weight individuals give each influence by creating a situation wherein these two influences come into conflict. While both factors are known to have an influence on individuals, the current study reveals their relative weighting. Results suggest that in-group bias has a stronger influence than attractiveness on decision-making.

A few factors could explain the findings. The context in which participants were required to decide may have led to these results. Based on previous research, the nonsignificant results suggest that although attractiveness has been shown to have a "halo effect" on individuals, its influence may not be as strong as in-group bias (Batres & Shiramizu, 2022). Such results highlight the importance of social context over individual traits. Another notable limitation is the lack of direct supervision during survey completion. Without supervision, there is a possibility of participants colluding with other participants. Such interactions could disrupt the Minimal Group Paradigm and affect the findings.

While it is clearly identified that in-group bias had a stronger influence than attractiveness bias, there was substantial variation in the responses of individuals. This suggests the possibility of other factors having an influence on the results. While this level of analysis is beyond the scope of the current study, a comprehensive analysis on some confounding variables is worthwhile for future research.

Individual Differences

Individual differences could have also played a role in the insignificant results. These differences affect how individuals interact with the world and could affect how participants perceive group dynamics and attractiveness (Newheiser et al., 2012). Although the relationship between attractiveness and answer change was insignificant, participants were most likely to retain their answers when they found the confederate unattractive. This suggests the possibility that the influence of attractiveness and in-group bias has differing levels of influence on different individuals.

This is not the first study identifying the impact of individual differences in terms of decision-making. A study looking at a dual-strategy model of reasoning suggests that people have two reasoning strategies which affect the way social cues affect them (Gagnon-St-Pierre et al., 2021). This includes a statistical strategy which involves estimating a likely conclusion of a social interaction or the counterexample strategies which involve generating counterexamples of a conclusion. Results from this study suggest that these processing distinctions underlie individual differences in responding to attractiveness and in-group bias.

Cultural Differences

The current study did not identify culture as a confounding variable, and this could prove to be a rich area of research for the future. The sample was a convenience sampling, meaning participants were selected due to availability. All participants were also enrolled in an Australian university. While their cultural identification was not collected, it can be assumed that students have had interactions with Australian culture, which is a highly individualistic culture (Emiko & Hidehumi, 2019). Cultural differences are known to affect how individuals perceive others who are part of their group (Sam & Berry et al., 2010). Hence, it is possible for individuals to have varying levels of loyalty to their group. Cox et al. (1991) examined the role of cultural backgrounds on teamwork. It was found that people from collectivist cultures displayed more cooperative behaviors compared to those from individualistic cultures. In other words, people from collectivist cultures place more emphasis on group success compared to individual success. In another study, Fischer and Derham (2016) examined cultural influences on in-group bias across 18 societies. The results showed that in-group bias was present but varied across the societies. The variation of in-group bias was dependent on the following cultural influences: individualistic versus collectivist cultures, uncertainty avoidance, and power distance. Similarly, it was found that collectivist cultures tend to exhibit stronger in-group biases. In cultures wherein hierarchical structures are more acceptable, more in-group behaviors were displayed. Cultures with high uncertainty avoidance, the extent to which individuals are affected by unfamiliar/unknown situations, displayed high levels of in-group bias. These cultural differences can affect how substantial the influence of attractiveness is in a group setting.

Personality Traits and Familiarity Bias

Personality traits have also been shown to influence responses to social cues. Traits such as agreeableness, extraversion, openness, and conscientiousness have been linked to group cohesion (Larsen et al., 2020; Saapna & Suman, 2012). This may shape how individuals engage with attractive others, affecting decision-making. Additionally, familiarity bias –where repeated exposure increases preference – was not accounted for in the current study (Monin, 2003). Prior research shows that even brief exposure to faces can enhance attractiveness ratings (Rhodes et al., 2001) and increased interaction fosters interpersonal attraction (Rei et al., 2001). Both these factors may have influenced responses and, therefore, should be considered as potential variables in future research examining the interplay between in-group bias and attractiveness.

Implications

The results from the current study suggest that facial attractiveness—a measure of overall attraction—is dependent on the context, and in situations where there are other social cues—such as in-group bias—it may not be as influential. This demonstrates the benefits of considering social context when looking at how attractiveness influences behaviors. Rather than solely addressing and mitigating the attractiveness bias in a group setting, it may be more advantageous to promote inclusive behaviors and encourage the acceptance of out-group individuals.

Results from previous experiments show that a properly designed Minimal Group Paradigm promotes in-group behaviors. This study's results support the fact that arbitrarily created groups can lead individuals to display in-group behaviors, as indicated by individuals choosing to change their answers when the group disagreed with them. These findings enhance the understanding of the underlying themes of group creation and allow for the promotion or mitigation of group biases.

Future Directions

Future research should consider replication of the current experiment in broader contexts with application of potential factors as described above and beyond,such as considering subjectivity due to race and/or ethnicity. Exploring dimensions such as personality traits and familiarity bias not only mitigates some individual differences but also offers a better understanding of different ways other dimensions interact with attractiveness and whether this could override other biases, such as in-group bias.

While the current study did not offer evidence for the effect of attractiveness on answer changes in group settings, it further supports evidence of in-group bias in decision-making. It enhances the understanding of social influences and encourages further research on how individual characteristics interact in group settings.

References

Adam-Troian, J., Bonetto, E., & De Oliveira, P. (2020). National identity salience increases prejudice: Evidence from a large-scale experiment. *Group Processes & Intergroup Relations, 23*(6), 849–866. https://doi.org/10.1177/1368430219876180

Batres, C., & Shiramizu, V. (2022). Examining the "attractiveness halo effect" across cultures. *Current Psychology, 42*(29), 25515–25519. https://doi.org/10.1007/s12144-022-03575-0

Baumeister, R. F., Ainsworth, S. E., & Vohs, K. D. (2015). Are groups more or less than the sum of their members? The moderating role of individual identification. *Behavioral and Brain Sciences, 39*. https://doi.org/10.1017/s0140525x15000618

Berscheid, E., Dion, K., Walster, E., & Walster, G. W. (1971). Physical attractiveness and dating choice: A test of the matching hypothesis. *Journal of Experimental Social Psychology, 7*(2), 173–189. https://doi.org/10.1016/0022-1031(71)90065-5

Calanchini, J., Schmidt, K., Sherman, J. W., & Klein, S. A. (2022). The contributions of positive outgroup and negative ingroup evaluation to implicit bias favoring outgroups. *Proceedings of the National Academy of Sciences, 119*(40). https://doi.org/10.1073/pnas.2116924119

Castelli, L., Tomelleri, S., & Zogmaister, C. (2008). Implicit ingroup metafavoritism: Subtle preference for ingroup members displaying ingroup bias. *Personality and Social Psychology Bulletin, 34*(6), 807–818. https://doi.org/10.1177/0146167208315210

Cellerino, A. (2003). Psychobiology of facial attractiveness. *Journal of Endocrinological Investigation, 26*(3), 45–48. https://pubmed.ncbi.nlm.nih.gov/12834020/

Chai, Q., Yin, J., He, J., & Lansu, T. A. M. (2022). Preschoolers' ingroup bias in predicting others' sharing: The role of contexts and theory of mind. *Journal of Experimental Child Psychology, 215*, 105340. https://doi.org/10.1016/j.jecp.2021.105340

Chen, F. F., & Kenrick, D. T. (2002). Repulsion or attraction?: Group membership and assumed attitude similarity. *Journal of Personality and Social Psychology, 83*(1), 111–125. https://doi.org/10.1037//0022-3514.83.1.111

Cox, T. H., Lobel, S. A., & McLeod, P. L. (1991). Effects of ethnic group cultural differences on cooperative and competitive behavior on a group task. *Academy of Management Journal, 34*(4), 827–847. https://doi.org/10.5465/256391

Cui, F., Deng, K., Liu, J., Huang, X., Yang, J., Luo, Y.-J., Feng, C., & Gu, R. (2023). Resource scarcity aggravates ingroup bias: Neural mechanisms and cross-scenario validation. *The British Journal of Psychology, 114*(4), 778–796. https://doi.org/10.1111/bjop.12654

Dang, J., Liu, L., Zhang, Q., & Li, C. (2019). Leaving an attacked group: Authoritative criticism decreases ingroup favoritism. *Journal of Pacific Rim Psychology, 13*. https://doi.org/10.1017/prp.2019.5

Downs, A. C., & Lyons, P. M. (1991). Natural observations of the links between attractiveness and initial legal judgments. *Personality and Social Psychology Bulletin, 17*(5), 541–547. https://doi.org/10.1177/0146167291175009

Espín, A. M., Correa, M., & Ruiz-Villaverde, A. (2019). Patience predicts cooperative synergy: The roles of ingroup bias and reciprocity. *Journal of Behavioral and Experimental Economics, 83*, 101465. https://doi.org/10.1016/j.socec.2019.101465

Fischer, R., & Derham, C. (2016). Is in-group bias culture-dependent? A meta-analysis across 18 societies. *SpringerPlus, 5*(1), 70–70. https://doi.org/10.1186/s40064-015-1663-6

Freud, E., Stajduhar, A., Rosenbaum, R. S., Avidan, G., & Ganel, T. (2020). The COVID-19 pandemic masks the way people perceive faces. *Scientific Reports, 10*(1), 22344–22344. https://doi.org/10.1038/s41598-020-78986-9

Gagnon-St-Pierre, É., Doucerain, M. M., & Markovits, H. (2021). Reasoning strategies explain individual differences in social reasoning. *Journal of Experimental Psychology. General, 150*(2), 340–353. https://doi.org/10.1037/xge0000852

Gangestad, S. W., & Scheyd, G. J. (2005). The evolution of human physical attractiveness. *Annual Review of Anthropology, 34*(1), 523–548. https://doi.org/10.1146/annurev.anthro.33.070203.143733

Griffin, A. M., & Langlois, J. H. (2006). Stereotype directionality and attractiveness stereotyping: Is beauty good or is ugly bad? *Social Cognition, 24*(2), 187–206. https://doi.org/10.1521/

soco.2006.24.2.187

Hare, B. (2017). Survival of the friendliest: Homo sapiens evolved via selection for prosociality. *Annual Review of Psychology, 68*(1), 155–186. https://doi.org/10.1146/annurev-psych-010416-044201

Horton, J., Rand, D., & Zeckhauser, R. (2010). The online laboratory: Conducting experiments in a real labor market. *National Bureau of Economic Research Working Paper Series, 15961*. https://doi.org/10.3386/w15961

Klebl, C., Rhee, J. J., Greenaway, K. H., Luo, Y., & Bastian, B. (2021). Beauty goes down to the core: Attractiveness biases moral character attributions. *Journal of Nonverbal Behavior, 46*(1), 83–97. https://doi.org/10.1007/s10919-021-00388-w

Kniffin, K. M., Wansink, B., Griskevicius, V., & Wilson, D. S. (2014). Beauty is in the in-group of the beholded: Intergroup differences in the perceived attractiveness of leaders. *The Leadership Quarterly, 25*(6), 1143–1153. https://doi.org/10.1016/j.leaqua.2014.09.001

Krendl, A. C., Magoon, N. S., Hull, J. G., & Heatherton, T. F. (2011). Judging a book by its cover: The differential impact of attractiveness on predicting one's acceptance to high- or low-status social groups. *Journal of Applied Social Psychology, 41*(10), 2538–2550. https://doi.org/10.1111/j.1559-1816.2011.00824.x

Kurzban, R., Tooby, J., & Cosmides, L. (2001). Can race be erased? Coalitional computation and social categorization. *Proceedings of the National Academy of Sciences, 98*(26), 15387–15392. https://doi.org/10.1073/pnas.251541498

Landy, J. F., Jia, M. (Liam), Ding, I. L., Viganola, D., Tierney, W., Dreber, A., Johannesson, M., Pfeiffer, T., Ebersole, C. R., Gronau, Q. F., Ly, A., van den Bergh, D., Marsman, M., Derks, K., Wagenmakers, E.-J., Proctor, A., Bartels, D. M., Bauman, C. W., Brady, W. J., ... Uhlmann, E. L. (2020). Crowdsourcing hypothesis tests: Making transparent how design choices shape research results. *Psychological Bulletin, 146*(5), 451–479. https://doi.org/10.1037/bul0000220

Larsen, M. M., Esenaliev, D., Brück, T., & Boehnke, K. (2020). The connection between social cohesion and personality: A multilevel study in the Kyrgyz Republic. *International Journal of Psychology, 55*(1), 42–51. https://doi.org/10.1002/ijop.12551

Li, W.-Q., Li, L. M., Jiang, D., & Liu, S. (2021). Fate control and ingroup bias in donation for the fight with the coronavirus pandemic: The mediating role of risk perception of COVID-19. *Personality and Individual Differences, 171*, 110456. https://doi.org/10.1016/j.paid.2020.110456

Lorenzo, G. L., Biesanz, J. C., & Human, L. J. (2010). What is beautiful is good and more accurately understood. *Psychological Science, 21*(12), 1777–1782. https://doi.org/10.1177/0956797610388048

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods, 47*(4), 1122–1135. https://doi.org/10.3758/s13428-014-0532-5

Ma, D. S., Kantner, J., & Wittenbrink, B. (2020). Chicago face database: Multiracial expansion. *Behavior Research Methods, 53*(3), 1289–1300. https://doi.org/10.3758/s13428-020-01482-5

Mackie, D. M., & Ahn, M. N. (1998). Ingroup and outgroup inferences: When ingroup bias overwhelms outcome bias. *European Journal of Social Psychology, 28*(3), 343–360. https://doi.org/10.1002/(sici)1099-0992(199805/06)28:3<343::aid-ejsp863>3.0.co;2-u

Marini, M., Ansani, A., Paglieri, F., Caruana, F., & Viola, M. (2021). The impact of facemasks on emotion recognition, trust attribution and re-identification. *Scientific Reports, 11*(1). https://doi.org/10.1038/s41598-021-84806-5

Maurer-Fazio, M., & Lei, L. (2014). 'As rare as a panda': How facial attractiveness, gender, and occupation affect interview callbacks at Chinese firms. SSRN. *Electronic Journal*. https://doi.org/10.2139/ssrn.2520780

Mcgarty, C., Haslam, S. A., Hutchinson, K. J., & Turner, J. C. (1994). The effects of salient group memberships on persuasion. *Small Group Research, 25*(2), 267–293. https://doi.org/10.1177/1046496494252007

Monin, B. (2003). The warm glow heuristic: When liking leads to familiarity. *Journal of Personality and Social Psychology, 85*(6), 1035–1048. https://doi.org/10.1037/0022-3514.85.6.1035

Montalan, B., Lelard, T., Godefroy, O., & Mouras, H. (2012). Behavioral investigation of the influence of social categorization on empathy for pain: A minimal group Paradigm Study. *Frontiers in Psychology, 3*. https://doi.org/10.3389/fpsyg.2012.00389

Newheiser, A.-K., & Dovidio, J. F. (2012). Individual differences and intergroup bias: Divergent dynamics associated with prejudice and stereotyping. *Personality and Individual Differences, 53*(1), 70–74. https://doi.org/10.1016/j.paid.2012.02.024

Nicholson, S. P., Coe, C. M., Emory, J., & Song, A. V. (2016). The politics of beauty: The effects of partisan bias on physical attractiveness. *Political Behavior, 38*(4), 883–898. https://doi.org/10.1007/s11109-016-9339-7

Nordholm, L. A. (1980). Beautiful patients are good patients: Evidence for the physical attractiveness stereotype in first impressions of patients. Social Science & Medicine. *Medical Psychology and Medical Sociology, 14*(1), 81–83. https://doi.org/10.1016/0160-7979(80)90013-2

Otten, S. (2016). The minimal group paradigm and its maximal impact in research on social categorization. *Current Opinion in Psychology, 11*, 85–89. https://doi.org/10.1016/j.copsyc.2016.06.010

Øvervoll, M., Schettino, I., Suzuki, H., Okubo, M., & Laeng, B. (2020). Correction: Filtered beauty in Oslo and Tokyo: A spatial frequency analysis of facial attractiveness. *PLOS ONE*, 15(2). https://doi.org/10.1371/journal.pone.0229523

Qualtrics. (2020). *Qualtrics XM (April/May 2024)*. Qualtrics. https://www.qualtrics.com

Rathbone, J. A., Cruwys, T., Stevens, M., Ferris, L. J., & Reynolds, K. J. (2023). The reciprocal relationship between social identity and adherence to group norms. *British Journal of Social Psychology, 62*(3), 1346–1362. https://doi.org/10.1111/bjso.12635

Rei, H. T., Maniaci, M. R., Caprariello, P. A., Eastwick, P. W., & Finkel, E. J. (2011). Familiarity does indeed promote attraction in live interaction. *Journal of Personality and Social Psychology, 101*(3), 557–570. https://doi.org/10.1037/a0022885

Reynolds, K. J., Turner, J. C., Alexander Haslam, S., Ryan, M. K., Bizumic, B., & Subasic, E. (2007). Does personality explain in-group identification and discrimination? Evidence from the minimal group paradigm. *British Journal of Social Psychology, 46*(3), 517–539. https://doi.org/10.1348/014466606x153080

Rudman, L. A., Ashmore, R. D., & Gary, M. L. (2001). "Unlearning" automatic biases: The malleability of implicit prejudice and stereotypes. *Journal of Personality and Social Psychology, 81*(5), 856–868. https://doi.org/10.1037//0022-3514.81.5.856

Sam, D. L., & Berry, J. W. (2010). Acculturation: When individuals and groups of different cultural backgrounds meet. *Perspectives on Psychological Science, 5*(4), 472–481. https://doi.org/10.1177/1745691610373075

Scheepers, D., Spears, R., Doosje, B., & Manstead, A. S. R. (2006). Diversity in in-group bias: Structural factors, situational features, and social functions. *Journal of Personality and Social Psychology, 90*(6), 944–960. https://doi.org/10.1037/0022-3514.90.6.944

Schitter, C., Fleiß, J., & Palan, S. (2019). To claim or not to claim: Anonymity, symmetric externalities and honesty. *Journal of Economic Psychology, 71*, 13–36. https://doi.org/10.1016/j.joep.2018.09.006

Shahani, C., Dipboye, R. L., & Gehrlein, T. M. (1993). Attractiveness bias in the interview: Exploring the boundaries of an effect. *Basic and Applied Social Psychology, 14*(3), 317–328. https://doi.org/10.1207/s15324834basp1403_5

Skinner, M., & Stephenson, G. M. (1981). The effects of intergroup comparison on the polarization of opinions. *Current Psychology, 1*(1), 49

Terry, D. J., Hogg, M. A., & McKimmie, B. M. (2000). Attitude-behaviour relations: The role of in-group norms and mode of behavioural decision-making. *British Journal of Social Psychology, 39*(3), 337–361. https://doi.org/10.1348/014466600164534

Tidwell, N. D., Eastwick, P. W., & Kim, A. (2017). She's not one of us: Group membership moderates the effect of fertility cues on attractiveness ratings. *Social Psychological and Personality Science, 8*(6), 689–697. https://doi.org/10.1177/1948550616676878

Tsukiura, T., & Cabeza, R. (2010). Shared brain activity for aesthetic and moral judgments: Implications for the beauty-IS-good stereotype. *Social Cognitive and Affective Neuroscience, 6*(1), 138–148. https://doi.org/10.1093/scan/nsq025

Voit, M., Weiß, M., & Hewig, J. (2021). The benefits of beauty–individual differences in the pro-attractiveness bias in social decision making. *Current Psychology, 42*(14), 11388–11402. https://doi.

org/10.1007/s12144-021-02366-3

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science, 17*(7), 592–598. https://doi.org/10.1111/j.1467-9280.2006.01750.x

Zebrowitz, L. A., & Franklin, R. G. (2014). The attractiveness halo effect and the babyface stereotype in older and younger adults: Similarities, own-age accentuation, and older adult positivity effects. *Experimental Aging Research, 40*(3), 375–393. https://doi.org/10.1080/0361073x.2014.897151

Zuckerman, M., Miyake, K., & Elkin, C. S. (1995). Effects of attractiveness and maturity of face and voice on interpersonal impressions. *Journal of Research in Personality, 29*(2), 253–272. https://doi.org/10.1006/jrpe.1995.1015

Table D1

Response Possibilities

Group A	Group B
Agree	Agree
	Disagree
Disagree	Agree
	Disagree

Note. This table displays all the possible responses from Group A and Group B.

Figure 1

Answer Change vs Attractiveness of Confederate

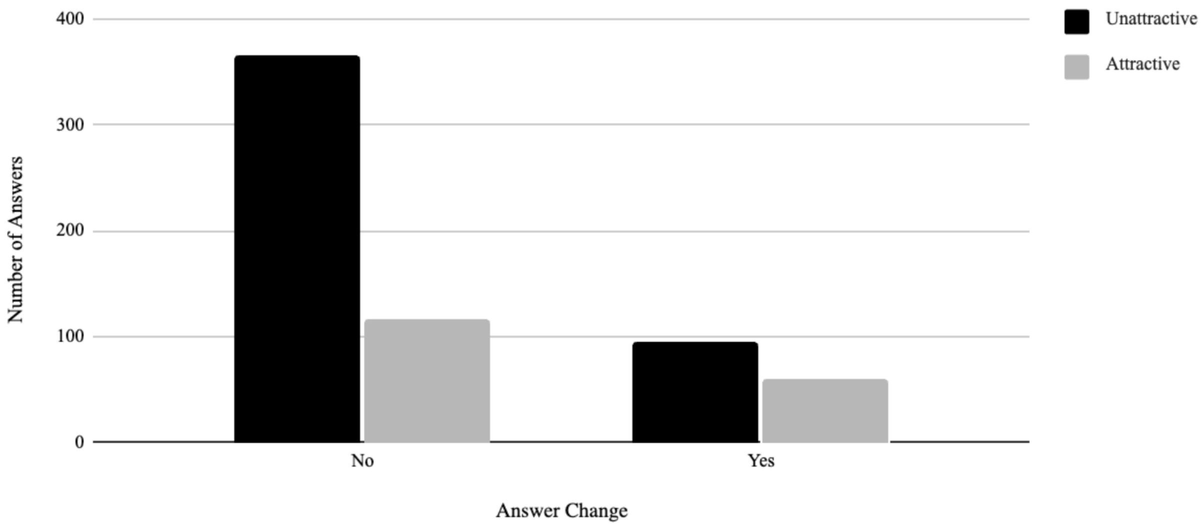


Figure 2

Group Response vs. Answer Change

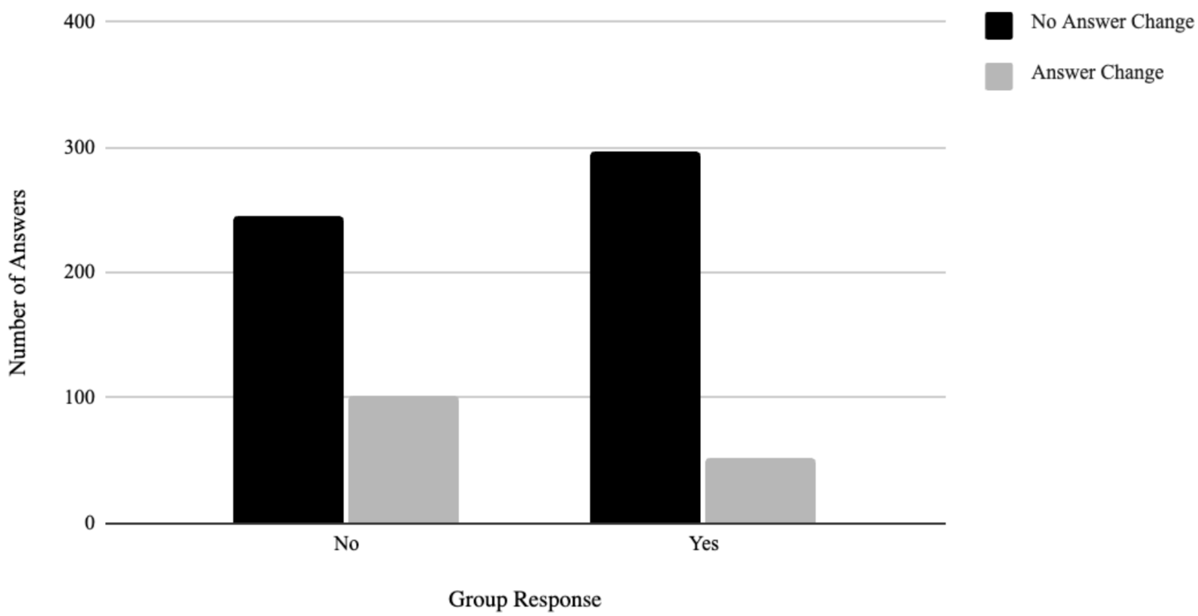


Figure 3

Attractiveness vs Answer Change when Group Agrees but Confederate Disagrees

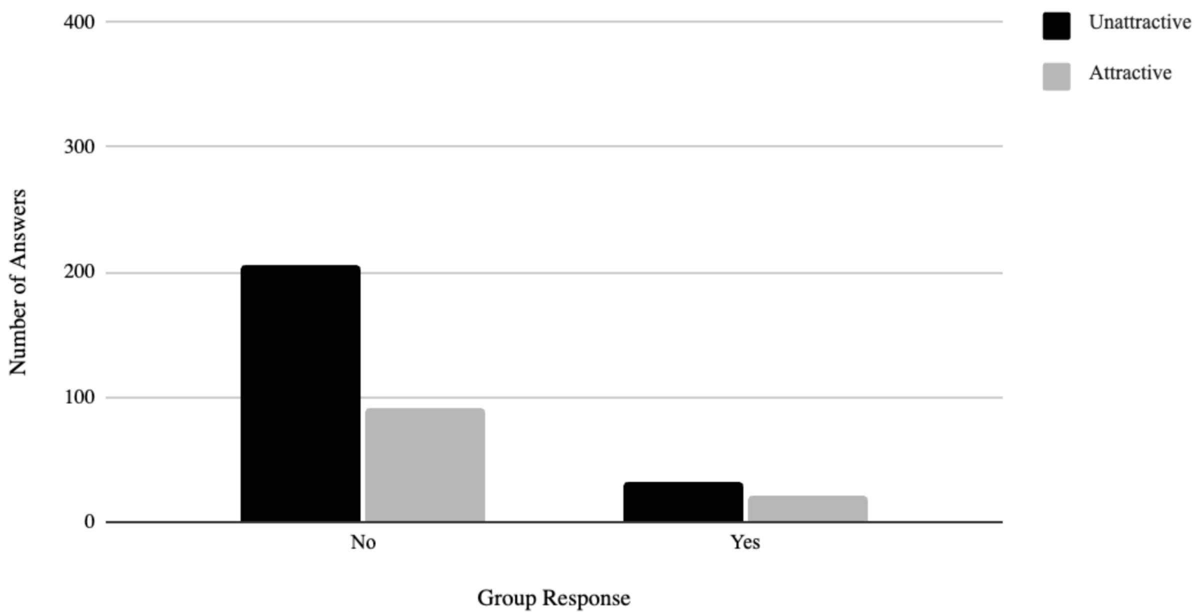
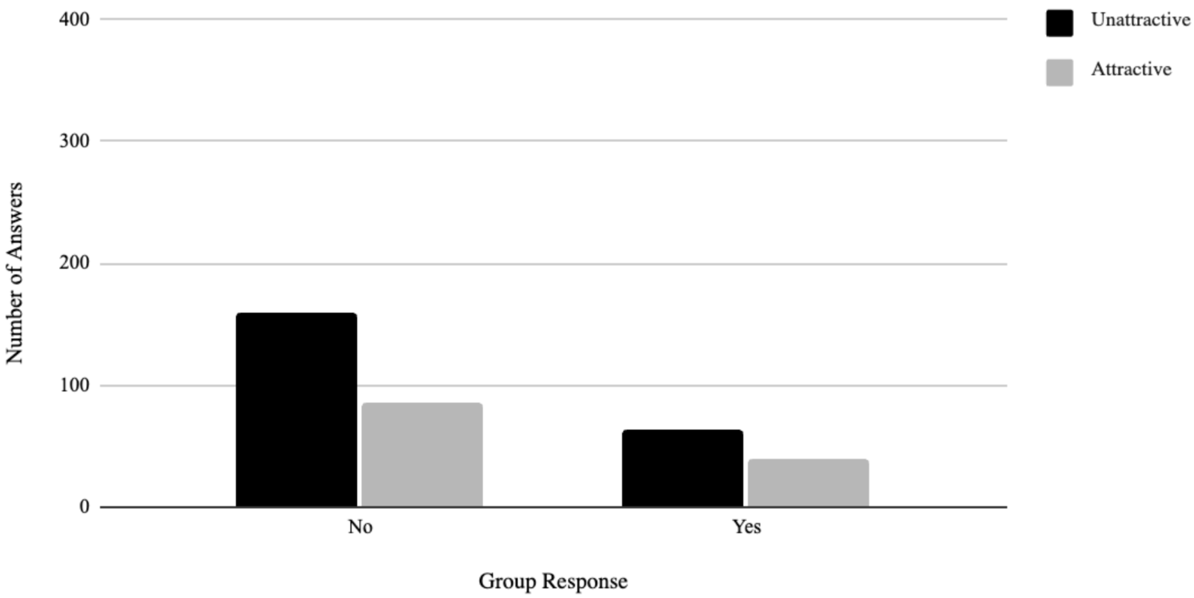


Figure 4

Answer Change vs. Attractiveness When Group Disagrees and Confederate Agrees



Appendix A

Quiz

This appendix consists of the mathematical questions and summaries of group and confederate answers.

1. $\sqrt{8.237}$ = ?
   a) 2.56
   b) 2.67
   c) 2.78
   d) 2.97

1a. Members of Group A agree with your answer.
   This member of group B disagrees with your answer.
   Are you going to change your answer?
      a) Yes
      b) No

2. $\sqrt{7.338}$ = ?
   a) 2.134
   b) 2.456
   c) 2.356
   d) 2.982

2a. Members of Group A disagree with your answer.
   This member of group B agrees with your answer.
   Are you going to change your answer?
      a) Yes
      b) No

3. $\sqrt{10.234}$ = ?
   a) 3.356
   b) 3.423
   c) 2.951
   d) 2.999

3a. Members of Group A agree with your answer.
This member of group B disagrees with your answer.
Are you going to change your answer?
a) Yes
b) No

4. $\sqrt{4.222}$ = ?
   a) 1.572
   b) 2.132
   c) 1.794
   d) 1.994

4a. Members of Group A disagree with your answer.
This member of group B agrees with your answer.
Are you going to change your answer?
a) Yes
b) No

5. √12.343 = ?
a) 2.988
b) 3.261
c) 3.942
d) 3.619

5a. Members of Group A agree with your answer.
This member of group B disagrees with your answer.
Are you going to change your answer?
a) Yes
b) No

6. √85.263 = ?
a) 9.163
b) 8.859
c) 7.295
d) 9.255

6a. Members of Group A disagree with your answer.
This member of group B agrees with your answer.
Are you going to change your answer?
a) Yes
b) No

7. √124.33 = ?
a) 12.483
b) 12.131
c) 11.173
d) 11.298

7a. Members of Group A agree with your answer.
This member of group B disagrees with your answer.
Are you going to change your answer?
a) Yes
b) No

8. √78.452 = ?
a) 7.762
b) 8.583
c) 7.589
d) 8.969

8a. Members of Group A disagree with your answer.
This member of group B agrees with your answer.
Are you going to change your answer?
a) Yes
b) No

Appendix B

Attractiveness Questionnaire

This appendix consists of the questions regarding whether participants find the confederates depicted attractive.

1. Is this person attractive?

A. Yes
B. No

2. Is this person attractive?

A. Yes
B. No

3. Is this person attractive?

A. Yes
B. No

4. Is this person attractive?

A. Yes
B. No

5. Is this person attractive?

A. Yes
B. No

6. Is this person attractive?

A. Yes
B. No

7. Is this person attractive?

A. Yes
B. No

8. Is this person attractive?

A. Yes
B. No

Appendix C

Contingency Tables

Table D2

Contingency Table for the Attractiveness of Confederate and Answer Change

Answer Change		Unattractive	Attractive	Total
No	Observed	365	177	542
	Expected	357	184.0	542
Yes	Observed	94	60	154
	Expected	102	52.40	154
Total	Observed	459	237	696
	Expected	459	237	696

Note. $n = 696$

Table D3

Contingency Table for Group Response vs Answer Change

Group Response		No Answer Change	Answer Change	Total
No	Observed	246	102	348
	Expected	271	77	348
Yes	Observed	296	52	348
	Expected	271	77	348
Total	Observed	542	154	696
	Expected	542	154	696

Note. $n = 696$

Appendix C (cont.)

Table D4

Attractiveness and Answer Change When the Group Agrees and Confederate Disagrees

Answer Change		Unattractive	Attractive	Total
No	Observed	205	91	236
	Expected	200.70	95.30	296
Yes	Observed	31	21	52
	Expected	35.3	16.7	52
Total	Observed	296	52	348
	Expected	296	52	348

Note. n = 348

Table D5

Attractiveness and Answer Change When the Group Agrees and Confederate Disagrees

Answer Change		Unattractive	Attractive	Total
No	Observed	160	86	246
	Expected	157.60	88.40	246
Yes	Observed	63	39	102
	Expected	65.40	36.60	102
Total	Observed	223	125	348
	Expected	223	125	348

Note. n = 348

FALL 2025 – VOLUME 25

Department of Counseling and Clinical Psychology
Teachers College, Columbia University
525 West 120th Street
New York, NY 10027
www.tc.edu/publications/gsjp

GRADUATE STUDENT JOURNAL of PSYCHOLOGY
EST 1998

Teachers College
COLUMBIA UNIVERSITY