Are Only the Unskilled Overconfident? Deconstructing the Dunning-Kruger Effect Through an Individual Differences Approach

Danielle N. Sanchez
University of New Mexico - Main Campus

Follow this and additional works at: https://digitalrepository.unm.edu/psy_etds

Part of the Cognitive Psychology Commons, and the Personality and Social Contexts Commons

Recommended Citation

This Thesis is brought to you for free and open access by the Electronic Theses and Dissertations at UNM Digital Repository. It has been accepted for inclusion in Psychology ETDs by an authorized administrator of UNM Digital Repository. For more information, please contact disc@unm.edu.
Danielle Sanchez-Combs
Candidate

Psychology
Department

This thesis is approved, and it is acceptable in quality and form for publication:

Approved by the Thesis Committee:

Dr. Eric Ruthruff, Ph.D., Chairperson

Dr. Tania Reynolds, Ph.D. Member, Thesis Committee

Dr. Ann Speed, Ph.D. Member, Thesis Committee
Are Only the Unskilled Overconfident? Deconstructing the Dunning-Kruger Effect

Through an Individual Differences Approach

by

Danielle Sanchez-Combs

B.S., Psychology, University of New Mexico, 2018

THESIS

Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Science

Psychology

The University of New Mexico Albuquerque, New Mexico

July 2021
Are Only the Unskilled Overconfident? Deconstructing the Dunning-Kruger Effect Through an Individual Differences Approach

by

Danielle Sanchez-Combs
B.S., Psychology, University of New Mexico, 2018
M.S., Psychology, University of New Mexico, 2021

ABSTRACT

This study sought to determine whether individuals suffering from the Dunning-Kruger effect are prone to misassessing their performance due to underlying personality and cognitive characteristics. To test this hypothesis, we first collected theory-informed measures of cognitive and personality traits. Next, we used three different performance estimate measures to assess the degree to which participants misestimated their abilities across two performance tasks (e.g., English grammar and logical reasoning). We found that some individuals are more prone to misassessing their performance and self-reported general Metacognitive Ability, Openness to Experience from the Big-Five personality Inventory, and an External Locus of Control orientation can play a role in this misestimation. In addition, individuals are also most likely to misassess their performance when comparing themselves to others or when evaluating their performance at the item-level than when assessing how many questions they will answer correctly.
TABLE OF CONTENTS

List of Figures .................................................................................................................. vi
List of Tables ..................................................................................................................... vii
I. Introduction .................................................................................................................... 1
   The Classic Dunning-Kruger Effect Paradigm ............................................................... 2
      Modification 1: Within-Subjects Design .................................................................... 4
      Modification 2: Pre- and Post-Task Measures of Performance ................................. 5
      Modification 3: A Third Measure of Performance Misestimation ............................ 5
      Modification 4: Measuring Individual Differences ...................................................... 7
   Using Foundational Theories to Identify Candidate Traits ........................................... 8
      Social Desirability ....................................................................................................... 11
      Dual-Burden: Low Skill Leads to Low Domain-Specific Metacognition ................. 12
      General Metacognitive Ability .................................................................................... 13
II. Research Predictions .................................................................................................... 19
   Prediction 1: Trait-Based Misestimation ....................................................................... 19
      Prediction 1.b.: Social Desirability Account .............................................................. 19
      Prediction 1.c.: Dual Burden Account ........................................................................ 19
      Prediction 1.d: Metacognitive Ability Account ......................................................... 19
III. Methods ....................................................................................................................... 21
   Participants ...................................................................................................................... 21
   Materials ......................................................................................................................... 23
      The Personality Battery .............................................................................................. 23
      The Performance Tasks ............................................................................................... 25
   Types of Performance Misestimation .......................................................................... 27
   Calculating Overall Scores and Outcome Measures for Plotting and Analysis .......... 27
IV. Results ......................................................................................................................... 30
   Participants’ Ability to Accurately Assess Their Performance ...................................... 30
   Overconfidence Patterns Across Domains .................................................................... 34
      Overplacing ................................................................................................................ 35
Global Overestimation ........................................................................................................... 36
Overprecision .............................................................................................................................. 37
Testing the Hypothesized Theoretical Pathways of Overconfidence................................. 38
Global Estimation and the Fit of the Foundational Theories ............................................... 40
Overplacing and the Fit of the Foundational Theories ......................................................... 44
Exploratory Analysis ............................................................................................................... 47
Discrimination and Sensitivity to Errors ............................................................................... 50

V. Discussion ............................................................................................................................... 55

Testing the Underlying Causal Theories of the Dunning-Kruger Effect ......................... 56
Fit of the Metacognitive Ability Account ............................................................................... 56
Fit of the Social Desirability Account .................................................................................... 58
Fit of the Dual-Burden Account ............................................................................................. 59

Alternative Theories of the Dunning-Kruger Effect ............................................................. 60

Research Limitations ............................................................................................................... 63
Summary .................................................................................................................................... 64

VI. References ............................................................................................................................. 66
LIST OF FIGURES

Figure 1. Novel Features of This Study.................................................................4
Figure 2. Implementation of Novel Approaches......................................................8
Figure 3. Theoretical Model of Performance Misestimation ....................................9
Figure 4. Mapping of Relationships Between Traits and Foundational Theories ........11
Figure 5. Overall Overplacing..............................................................................31
Figure 6. Overall Global Overestimation...............................................................32
Figure 7. Overall Overprecision......................................................................33
Figure 8. Average Confidence on Error and Correct Trials by Quartile ..................51
Figure 9. Example Calibration Curve from Lichtenstein, Fischhoff, & Philips (1976) ....53
Figure 10. Calibration Curve Analysis.................................................................54
LIST OF TABLES

Table 1. Employment Status .................................................................................................................. 22
Table 2. Income ........................................................................................................................................ 23
Table 3. Ethnicity ...................................................................................................................................... 23
Table 4. Measures of Personality and Cognition ..................................................................................... 24
Table 5. How Estimates and Accuracy were Aggregated for Plotting and Analysis .............................. 28
Table 6. Mean and Standard Deviations for Outcome Measures ............................................................. 29
Table 7. Regression Results Using Estimated Grammar Percentile as the Criterion .............................. 35
Table 8. Regression Results Using Grammar Global Overestimation as the Criterion ............................. 36
Table 9. Regression Results Using Grammar Overprecision as the Criterion ......................................... 37
Table 10. Means, Standard Deviations and Correlations between Dependent and Independent Variables ................................................................................................................................................. 39
Table 11. Results of Hierarchical Regression Analysis for the Metacognitive Pathway and Global Estimation .............................................................................................................................................................................. 40
Table 12. Results of Hierarchical Regression Analysis for the Social Desirability Pathway and Global Estimation .............................................................................................................................................................................. 42
Table 13. Results of Hierarchical Regression Analysis for the Metacognitive Pathway and Overplacing .............................................................................................................................................................................. 44
Table 14. Results of Hierarchical Regression Analysis for the Social Desirability Pathway and Global Estimation .............................................................................................................................................................................. 45
Table 15. Results of Hierarchical Regression Analysis for Overall Global Estimation ............................. 48
Table 16. Results of Hierarchical Regression Analysis for Overall Overplacing ........................................ 49
I. Introduction

From engineering to medicine and beyond, we rely on experts to accurately assess their capability to perform their jobs. Accurate judgment requires an individual to understand when they know an answer and when they do not. Yet, research shows that a high percentage of individuals are overly optimistic about their performance (Kruger & Dunning, 1999; Ehrlinger et al., 2008; Dunning et al., 2003; Moore, 2008). Those who are the most optimistic regarding their abilities also tend to be the lowest performers. On the other hand, high performers tend to suffer from the opposite performance-estimation error - they consistently underestimate their performance. This differential pattern of lack of self-calibration between low and high performers is known as the Dunning-Kruger effect (Kruger & Dunning, 1999; Dunning et al., 2003).

The Dunning-Kruger effect is a far-reaching phenomenon that has been found in both high-consequence situations and everyday matters. For instance, the Dunning-Kruger effect has been found to have occurred during the 9/11 and Hurricane Katrina emergency evacuations (Siems, 2016), in gun safety knowledge examination scores (Ehrlinger et al., 2008), in the endorsement of anti-vaccine policy attitudes (Motta, Callaghan, & Sylvester, 2018), and in misdiagnosing brain death (Redelmeier & Scales, 2015). The Dunning-Kruger effect has also been found in more basic domains such as English grammar, humor, logical reasoning (Kruger & Dunning, 1999; Pennycook et al., 2017), and financial literacy (Sanchez & Dunning, 2018).

Despite the large body of literature on the Dunning-Kruger effect, individual differences in the expression of the Dunning-Kruger effect remain essentially
unexplored. Without a full understanding of the role that cognitive and personality traits play in the Dunning-Kruger effect, the field remains stunted in its ability to accurately predict who is most likely to over or underestimate their performance and how best to correct for it. This can have severe consequences when accuracy matters most, such as when diagnosing brain death or conducting emergency evacuations.

Therefore, the goal of the present research is to determine whether cognitive and personality traits make some individuals more prone to misestimating their performance. Dunning-Kruger research often attributes performance misestimation to task skill (Kruger & Dunning, 1999; Dunning, 2011). While task skill is important to know, it is difficult to draw predictions beyond the particular task that is being studied. In contrast, having a “profile” of who is most likely to misestimate their performance based on individual characteristics will allow for better prediction of misestimation across domains.

The rest of the Introduction is organized as follows. First, I will describe the typical Dunning-Kruger paradigm and the innovations that were brought to this classic experimental set-up to enable an individual differences investigation. Second, I will discuss the foundational theories used to explain the cause of the Dunning-Kruger effect and how these theories enabled the identification of pertinent individual traits.

The Classic Dunning-Kruger Effect Paradigm

In their seminal work, Kruger and Dunning (1999) produced a paradigm that has formed the basis of many Dunning-Kruger investigations since. In their original set of studies, Kruger and Dunning investigated participants’ awareness of their ability to correctly identify humorous material (Study 1), to think logically and analytically on a
set of logical reasoning problems (Study 2), and to know when they have correctly identified and corrected ungrammatical English sentences and passages (Study 3).

Their experimental design was relatively simple. Participants were given a set of either 20 questions from a logical reasoning or English grammar task or a set of 30 questions from a humor task. After completing the task, they were asked to provide three estimates about their performance. First, participants were asked to compare their general ability with that of other students in their course on a percentile scale. Second, they were asked to estimate how their score would compare to that of other students, also on a percentile scale. Finally, they were asked to estimate how many questions they had gotten correct on the task.

Kruger and Dunning (1999) found that participants who scored poorly on the task greatly overestimated how their performance and score compared to others, as well as overestimated how many questions they had answered correctly. However, the top scorers underestimated their performance and score compared to others but did not significantly underestimate the number of questions they had answered correctly.

While this initial formulation of the Dunning-Kruger paradigm has been repeated many times, producing robust effects, it unfortunately does not properly lend itself to an individual differences investigation. The present study has therefore implemented several modifications to the original Dunning-Kruger paradigm to better capture individual differences in performance misestimation. These modifications are shown in Figure 1.
Novel Features of This Study

Comparing Types of Performance Misestimation

- Percentile Ranking
- Item-by-Item Estimates
- Global Estimates

Capturing Estimates of Performance at Different Time Points

- Pre-Task Estimates
- During-Task Estimates
- Post-Task Estimates

Tools to Build a Profile of Misestimators

- Personality & Cognitive Measures to Predict Misestimation
- Use of Two Tasks: Grammar & Logic

Note. This study collected three types of performance estimates to clarify what type of errors participants were making when misassessing their performance. Performance estimates were also collected pre-, during-, and post-task, to compare baseline estimates of performance to estimates after exposure to each task. To understand who is most likely to misassess their performance, I collected several personality and cognitive trait measures and had participants complete a logic and grammar task. By using two tasks that are not strongly correlated with one another, I could better examine if performance misestimation on one task predicted performance misestimation on the other task.

Modification 1: Within-Subjects Design

The original Dunning-Kruger paradigm gave the humor, logical reasoning, and grammar tests to separate samples of participants, which precludes the ability to see whether some participants systematically misassess their performance, regardless of the task at hand. Therefore, this study used a within-subjects design in which all participants received both a logical reasoning and an English grammar task. By using two relatively unrelated tasks within the same sample, I could identify whether misestimation in one domain carries over to misestimation in another. I could then examine the power of personality and cognitive measures to predict overestimation across domains.
Modification 2: Pre- and Post-Task Measures of Performance

A second modification implemented in this study was asking participants to estimate their performance before and after completing the tasks, which is typically not done in Dunning-Kruger effect studies. Capturing pre- and post-estimates of performance offered two advantages: 1) the pre-task estimates highlighted participants’ view of their baseline abilities and 2) by having both pre- and post-task estimates, I could see how performance estimates changed after exposure to the experimental tasks.

If there is a large degree of difference between participants’ pre-estimates (e.g., their baseline belief in their abilities) and their post-estimates, then this would indicate that task-specific factors, like task difficulty, alters participants’ perception of their performance. The influences of task specific factors, like task difficulty, are thought to have a large impact on performance estimations (e.g., Burson, Larrick, & Klayman, 2006; Krueger & Mueller, 2002) though this finding is debated by others in the field (Ehrlinger et al., 2008). If there is little adjustment in individuals’ pre- and post-estimates, it could indicate that prior familiarity with the task domains is directly influencing the production of faulty performance expectations.

Modification 3: A Third Measure of Performance Misestimation

Past research on the Dunning-Kruger effect typically uses two methods of capturing under and overestimation: 1) asking participants how many questions they answered correctly at the end of the task (i.e., global accuracy estimates) and 2) asking participants to estimate their achieved percentile (i.e., percentile ranking estimates) (Kruger & Dunning, 1999; Ehrlinger et al., 2008; Pavel et al., 2012; Pennycook et al.,
However, global accuracy and percentile estimates are prone to forgetting, as they require individuals to keep a running-tally of how many questions were answered correctly. In addition, global accuracy estimates may be solely derived from participants’ base rates\(^1\), without participants taking into account task-specific factors like level of difficulty. Percentile rankings, on the other hand, require that participants not only assess their performance, but also that of others. Therefore, percentile rankings conflate two sources of misestimation (i.e., oneself and others).

To circumvent the potential problems with the traditional measures in this paradigm (i.e., percentile and global accuracy estimates), I also implemented a third type of performance estimate: item-by-item estimates of accuracy. Item-by-item estimates ask participants to estimate the likelihood that each answer they provide is correct, which may provide more accurate assessments than either global accuracy estimates or percentile rankings. Item-by-item assessments are often used in the metacognitive literature to assess judgments but have been used in only a single Dunning-Kruger study (e.g., McIntosh et al., 2019).

Item-by-item estimates are also captured in real-time, which can counteract forgetting effects associated with capturing estimates after the task is completed. Item-by-item estimates may also be less prone to problems with base rates, as participants are less likely to have available base rates for each individual question. Item-by-item estimates also isolate participants’ misunderstanding of their own performance from base rates.

---

1 Base rates are participants’ knowledge of their past performance, which has been derived from prior experiences, feedback, and/or grades. The problem with using base rates alone to estimate performance is that participants may be more inaccurate in assessing their performance because they are not taking into account task-specific factors like level of difficulty. It is akin to a participant saying they’re most likely to get a B on an exam because they typically receive B’s on their tests, regardless of how easy or hard the new test is.
that of misunderstanding others’ performance, which is the problem with using percentile rankings alone. Finally, item-by-item measures also allowed me to calculate participants’ ability to detect when they are wrong versus when they are right (e.g., sensitivity) and how well their estimates of accuracy match their actual performance (e.g., calibration; McIntosh et al., 2019).

**Modification 4: Measuring Individual Differences**

The final modification, and the most critical for the present study, is the addition of measures of individual differences in personality and cognitive style. This modification allowed me to determine which traits are predictive of over or underestimating one’s performance across domains. Because previous Dunning-Kruger studies have not included individual difference measures, it was not obvious which ones to use. In this regard, the present study is largely exploratory. In the next section, however, I will review the foundational theories regarding why individuals are over or underconfident in their performance and use these foundational theories to identify candidate traits that could cause misestimation. The modifications discussed above produced the overall experimental design shown in Figure 2 below.
Note. Before starting the task, participants were asked to provide estimates of their performance. Following this, they completed several personality and cognitive measures. They then completed a logical reasoning and English grammar task, all while provided confidence judgments regarding their accuracy for each question. Finally, they provided estimates of their performance at the end of the task.

Using Foundational Theories to Identify Candidate Traits

To identify personality and cognitive traits of interest, I leveraged three causal theories of performance misestimation from the Overconfidence and Dunning-Kruger effect literatures. Figure 3 is a visual representation of these three causal theories. The figure starts with Overconfidence, which is an umbrella term for misestimating one’s performance consistent with Moore’s (2008) naming convention.
Note. The literature on performance misestimation proposes three different mechanisms of overconfidence in one’s abilities. The three theories can be broken down into “deliberate misestimation” and “unintentional misestimation” pathways. Under the deliberate misestimation pathway, participants who score high on measures of Social Desirability may strategically overinflate their estimates to appear in a better light. Under the unintentional misestimation pathways two theories fall: 1) Dunning-Kruger’s theory of low task skill causing poor domain specific metacognition and 2) the theory of low general metacognitive ability, which is the inability to accurately understand one’s thinking, regardless of task experience or skill.

Theoretically, performance misestimation could follow two paths: 1) the “deliberate misestimation” path, in which people knowingly choose to moderate their estimates to appear in a better light (Bensch et al., 2019) or 2) the “unintentional misestimation” path, in which individuals are unaware that they are providing inaccurate self-assessments (Kruger & Dunning, 1999; Dunning, 2011).

Knowingly misestimating one’s performance can be caused by Social Desirability, which is a trait in which people seek social approval by acting in a manner that is socially acceptable (Pedregon et al., 2012). Social Desirability can cause people to increase their performance estimates to seem more knowledgeable or to decrease their estimates to appear humble (Pedregon et al., 2012).
Unknowingly misestimating one’s performance could be caused by two mechanisms. The first mechanism is poor domain-skill, which prevents individuals from understanding what good performance should look like and causes misestimation (Dunning, 2011). The second mechanism is poor Metacognitive ability, which harms an individual’s ability to understand their cognition, knowledge, and task performance across many domains, despite how skilled they may be.

In this individual differences investigation, I first measured the core theoretical constructs of Social Desirability, Metacognitive ability, and skill by using well-validated measures. I then used these core theories to identify additional traits that could impact performance misestimation. These additional traits are correlates of the core constructs, which had the added benefit of clarifying which aspects of these core theories are critical to inducing performance misestimation. Figure 4 below shows the full mapping of traits that were used in this study.
Figure 4

Mapping of Relationships Between Traits and Foundational Theories

Note. The figure above shows the relationship between traits. The blue box represents performance misestimation, which is the outcome measure of this study. The black boxes with the red numbered circles represent foundational theories that are existing causal explanations for why some individuals misassess their performance. These foundational theories include Social Desirability, Metacognition, and skill causing low domain-specific metacognition. The purple boxes represent traits that have correlated with Social Desirability, while the green boxes represent traits that have correlated with Metacognition.

Social Desirability

Individuals high in Social Desirability may have some sense of their true performance. However, those who perform poorly on a task may choose to purposefully inflate their performance estimates to appear more knowledgeable than they are. Those who perform well may purposefully underestimate their performance to appear humbler.

Three additional traits, Agreeableness, Extraversion, and overclaiming of knowledge may influence the relationship between Social Desirability and
misestimation. Individuals who score high on the Big-Five dimension of Agreeableness and Extraversion tend to desire positive social evaluations (Ashton and Lee, 2001; Chiaburu et al., 2015; Wilt and Revelle, 2009) and have previously been found to be overconfident in their abilities (Sukenik, Reizer, & Koslovsky, 2018). In addition, research has found a link between Extraversion and greater overconfidence in older adult populations (Dahl et al., 2010).

Socially desirable responding has also been positively correlated with overclaiming of knowledge. Overclaiming of knowledge is defined as claiming unfounded knowledge about a person, place, or thing that does not exist (Paulhus et al., 2003). This correlation may indicate that individuals who care about appearing in a positive light to others may also claim knowledge of fake terms for the same reason, to appear more knowledgeable (Paulhus & John, 1998).

Overclaiming is typically measured by Paulhus et al.’s (2003) Overclaiming Questionnaire. Examples from the Overclaiming questionnaire include a historical figure named “Queen Shattuck” and a literary term “sentence stigma”. These are not real people or concepts, but some people will nevertheless rate them as highly familiar (Paulhus et al., 2003). Overclaiming of one’s knowledge has also been correlated with the Dunning-Kruger effect in a previous research study examining fake news detection (Pennycook & Rand, 2018).

**Dual-Burden: Low Skill Leads to Low Domain-Specific Metacognition**

The theory developed by Kruger and Dunning (1999) to explain performance misestimation states that people who overestimate their performance do so because they lack domain-specific skill. Being unskilled is said to create a “dual-burden,” as the
ability to evaluate whether a response is correct relies on the skill to produce a correct response in the first place (e.g., domain-specific metacognition). Skill in the Dunning-Kruger effect is typically measured by overall task score (e.g., proportion correct).

In the Kruger and Dunning view, there is little room for individual differences other than skill. Therefore, when skill improves, so should performance estimates (Kruger & Dunning, 1999). However, this dual-burden account does not explain the underestimation seen in top performers. To remedy this problem, Dunning (2011) therefore added a social comparison account. In Dunning’s view, top performers hold the perception that others are just as knowledgeable and skilled as they are, and therefore underrate their performance. This implies that it is not skill that causes this misestimation, but a faulty worldview of how others would perform.

While the dual-burden explanation is certainly plausible, it does not explain why top performers underestimate their performance on measures that do not require them to compare their performance to others. In Ehrlinger et al. (2008), participants were asked to estimate how many questions they had answered correctly at the end of a gun knowledge assessment. Underestimation was still seen in top performers, despite not requiring them to compare themselves to others. Therefore, it is likely that other mechanisms impact assessment errors in top performers.

**General Metacognitive Ability**

While Kruger and Dunning attribute over and underestimation to domain-specific metacognition, tied directly to skill within that domain/task, they do not propose any role for overall (domain-general) metacognitive ability. Metacognition can be defined as an individual’s understanding of their own thinking processes (Lai, 2011).
Metacognitive ability allows an individual to understand their cognition and knowledge, deploy cognitive strategies, evaluate their task performance, and perceive how environmental and personal factors could impact current and future performance (Lai, 2011). Baseline metacognitive ability has been shown to vary widely from person to person (Kelemen, Frost, & Weaver III, 2000) and would likely affect performance across many domains, not just those in which individuals are unskilled. Therefore, I am uncoupling Metacognitive ability from task skill to determine the unique role that each play in performance misestimation.

Prior research has shown that Metacognitive ability can be essential to producing more accurate confidence judgments. A study conducted by Buratti, Allwood, and Kleitman (2013) explored the link between metacognitive ability, confidence levels, personal characteristics, and individuals’ ability to successfully recalibrate their confidence levels upon further reflection. In this investigation, participants were given a number of individual difference measures, that captured features such as self-doubt, Narcissism, the Big-Five personality traits, and how much individuals enjoyed engaging in thought (i.e., Need for Cognition). They found that an Openness factor, which was comprised of Openness to Experience from the Big-Five and Need for Cognition, predicted overconfidence on first and second-order confidence judgments, when controlling for score on the task. Those who were high in the Extraversion/Narcissism factor were found to be more confident on their correct trials, when controlling for participants’ score on the task.

---

2 The Big-Five personality traits consists of Conscientiousness, Openness to Experience, Neuroticism, Agreeableness, and Extraversion. Please see McCrae, Costa, & Martin (2005) for review.
While Buratti, Allwood, and Kleitman's investigation is similar in nature to the current study, it deviates in two critical ways. First, the goal of their research was to understand which traits would help individuals successfully recalibrate their confidence, not to determine if some individuals are more prone to over/underconfidence (e.g., searching for the existence of an over/underconfidence trait in the Dunning-Kruger effect). Second, it is not looking across types of confidence errors, such as social comparison (i.e., percentile rankings), overall judgments of accuracy (i.e., global accuracy estimates), and item-by-item estimates. Their study utilized item-by-item judgments alone to assess first-order and second-order confidence judgments.

Beyond the traits of Openness and Extraversion/Narcissism, prior research has shown that a number of additional traits correlate with Metacognitive ability. Throughout the next several sections I will examine each of these traits and how they could potentially influence an individual to over or underestimate their performance.

**Metacognition: The Big-Five Personality Traits.** People with high levels of conscientiousness have been found to score higher on measures of metacognition (Kelly & Donaldson, 2016; Chiaburu, Cho, & Gardner, 2015b; Winne, 1996; Buratti, Allwood, & Kleitman, 2013; Pennycook et al., 2017) and to score lower on overconfidence measures. Those who are high in conscientiousness are more likely to devise optimal learning and/or work strategies (Barrick and Mount, 1991; Judge and Ilies, 2002). The development of optimal strategies and the ability to recognize when a strategy is not working is dependent on the Metacognitive ability to correctly evaluate one’s performance (Lai, 2011).
A metacognition study examining participants’ confidence in their initial and second-order answers\(^3\) have also found a connection between Openness to Experience, Extraversion, and confidence ratings (Buratti, Allwood, & Kleitman, 2013). Individuals who scored high in Openness to Experience were found to be more confident on their incorrect trials than their correct trials, while those who were high in Extraversion were found to be more confident on their correct trials when controlling for participants’ score on the task.

**Metacognition: Feeling of Knowing.** Metacognition is often broken down into two categories: knowledge and regulation (Lai, 2011). Metacognitive knowledge includes knowledge about oneself and knowledge of factors that might impact performance. Metacognitive regulation is the monitoring of one’s thoughts and awareness of comprehension and task performance (Lai, 2011). Related to the knowledge aspect of metacognition is *Metamemory*, which is knowing what one knows and being able to recall it (Schwartz & Metcalfe, 2017). However, failures in Metamemory have been linked to overconfidence through false *Feelings of Knowing*. Feeling of Knowing is a top-down assessment of whether knowledge is within one’s experience and that an answer can be derived from one’s memory (Reder & Ritter, 1992; Atir, Rosenzweig, & Dunning, 2015).

False Feelings of Knowing can be hidden from individuals during the problem-solving process. When people first attempt to answer a question, they use metacognitive processes to decide whether they 1) already know the answer, 2) must deduce the answer, 3) must seek out additional information, or 4) could never answer

---

\(^3\) Second-order answers indicate answer changes.
the question because it is unanswerable (Reder & Ritter, 1992). When unsure of the answer, individuals may unconsciously infer answers that appear reasonable or expected (Dunning, 2011). For example, if you asked an individual whether Salem, Washington is a valid city and state pairing, many individuals may say that this is a valid pairing, because Salem is a city in the northwestern United States and Washington is a state within that same geographical region. However, Salem, Oregon is the correct city and state pairing. Because these inferential processes are mostly unconscious, the individual remains unaware of the fact that they have produced an educated guess rather than directly retrieving actual knowledge from memory (Dunning, 2011). Individuals may get a “gut feeling” or sense that Salem, Washington is a correct pairing because it *feels familiar*.

False Feelings of Knowing has also been linked to individual traits. A bias towards false Feelings of Knowing has been linked to overclaiming of knowledge (Goecke et al., 2020) and having an intuitive decision style (Sauter, 1999; Epstein et al., 1996). An intuitive decision style is the tendency to make decisions based on what feels right (Leykin & DeRubeis, 2010). Pennycook et al. (2017) has found that those with an intuitive decision style were more likely to overestimate their performance on a logical reasoning task, while those who had an analytical decision style were more likely to underestimate their performance. Individuals with analytical decision styles have also been found to have higher metacognitive ability (Petty et al., 2009). Additionally, intuitive decision style has been correlated with external locus of control. External locus of control is the belief that outside forces dictate the outcomes in one’s life (Fletcher et al., 1986). Therefore, individuals who are more overconfident in their abilities may be
more susceptible to false feelings of knowing, have an intuitive decision style, and an external locus of control orientation.
II. Research Predictions

**Prediction 1: Trait-Based Misestimation.** Some individuals are prone to over or underestimating their performance due to underlying cognitive and/or personality traits. If this is the case, individuals who over or underestimate should do so across tasks, regardless of skill level. This performance misestimation will be predicted by individual traits.

**Prediction 1.b.: Social Desirability Account.** If individuals are aware of their true performance but choose to inflate their estimates to appear in a better light, than overestimation should be predicted by Social Desirability scores and correlates of Social Desirability, which include overclaiming of one’s knowledge, Agreeableness, and Extraversion.

**Prediction 1.c.: Dual Burden Account.** If task skill directly determines over and underconfidence (as in the dual-burden account of Kruger & Dunning, 1999), than participants should be overconfident in domains in which they are low skilled and underconfident in domains in which they are high skilled. Over and underconfidence will not be predicted by individual traits.

**Prediction 1.d: Metacognitive Ability Account.** If overestimation is due to poor Metacognitive ability, then measures of metacognition should predict misestimation across domains, when controlling for task skill (as measured by task performance). The following correlates of Metacognitive ability may also be predictive of misestimation, which will allow us to refine the Metacognitive ability account.

- Intuitive decision style
- Overclaiming one’s knowledge
- High Openness to Experience
- Low conscientiousness
- An external locus of control orientation
III. Methods

Participants

This study recruited students enrolled in Psychology classes at the University of New Mexico (UNM) who were able to read and speak fluent English and were 18 years of age or older. Recruitment was done via email and sending flyers to course instructors. All participants completed the study online via Survey Monkey, after signing up via UNM’s SONA research system.

A statistical power analysis was conducted in order to determine how many participants would be necessary to run planned linear regression analyses between overestimation measures and individual differences. 293 participants took part in this study. A power analysis indicated a sample size of 180 would provide 0.95 power to detect a Cohen’s $f^2$ correlation of .35 (large) at an alpha of .05 for the planned linear regression analysis (Faul, Erdfelder, Lang, & Buchner, 2007).

Three metrics were used to determine whether a participant should be excluded for not following study protocols. Participants were removed if they did not complete both experimental tasks in their entirety, if they spent less than 30 minutes completing all 323 task and personality questions, and if they had failed one or more of the attention checks questions. Of the 293 participants, 25 were removed from the analysis based on these criteria. As our planned linear regression analyses examining individual differences used age and gender as control variables, we dropped an additional two participants who did not specify their gender. Overconfidence is known to increase with age (Menkhoff et al., 2013) and men are often found to be more overconfident than women (Pulford & Colman, 1997).
This left a total of 266 participants (197 females, 69 males) ranging from 18 to 57 years of age (M = 22.7, SD = 7.36). In addition to age and gender, I also collected employment status, yearly income, and ethnicity demographics. The majority of our sample listed their employment status as student (44.0%) or employed for wages (36.5%), made less than $20,000 per year (68.8%), and were of Hispanic or Latino (41%) or White/Caucasian (40.2%) descent. The results of these demographic questions can be found in Table 1, Table 2, and Table 3.

Table 1

*Employment Status*

<table>
<thead>
<tr>
<th>Levels</th>
<th>Counts</th>
<th>% of Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prefer Not to Answer</td>
<td>4</td>
<td>1.5%</td>
</tr>
<tr>
<td>Employed for Wages</td>
<td>97</td>
<td>36.5%</td>
</tr>
<tr>
<td>Self-Employed</td>
<td>4</td>
<td>1.5%</td>
</tr>
<tr>
<td>Out of Work, but Looking</td>
<td>16</td>
<td>6.0%</td>
</tr>
<tr>
<td>Out of work, but Not Looking</td>
<td>15</td>
<td>5.6%</td>
</tr>
<tr>
<td>Homemaker</td>
<td>6</td>
<td>2.3%</td>
</tr>
<tr>
<td>Student</td>
<td>117</td>
<td>44.0%</td>
</tr>
<tr>
<td>Military</td>
<td>3</td>
<td>1.1%</td>
</tr>
<tr>
<td>Retired</td>
<td>2</td>
<td>0.8%</td>
</tr>
<tr>
<td>Unable to Work</td>
<td>2</td>
<td>0.8%</td>
</tr>
</tbody>
</table>
Table 2

*Income*

<table>
<thead>
<tr>
<th>Levels</th>
<th>Counts</th>
<th>% of Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prefer Not to Answer</td>
<td>32</td>
<td>12.0%</td>
</tr>
<tr>
<td>Under 20K</td>
<td>183</td>
<td>68.8%</td>
</tr>
<tr>
<td>20-40K</td>
<td>30</td>
<td>11.3%</td>
</tr>
<tr>
<td>40-60K</td>
<td>7</td>
<td>2.6%</td>
</tr>
<tr>
<td>60-80K</td>
<td>5</td>
<td>1.9%</td>
</tr>
<tr>
<td>80-100K</td>
<td>4</td>
<td>1.5%</td>
</tr>
<tr>
<td>100K or Over</td>
<td>5</td>
<td>1.9%</td>
</tr>
</tbody>
</table>

Table 3

*Ethnicity*

<table>
<thead>
<tr>
<th>Levels</th>
<th>Counts</th>
<th>% of Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prefer Not to Answer</td>
<td>4</td>
<td>1.5%</td>
</tr>
<tr>
<td>Asian / Pacific Islander</td>
<td>24</td>
<td>9.0%</td>
</tr>
<tr>
<td>Black or African American</td>
<td>3</td>
<td>1.1%</td>
</tr>
<tr>
<td>Hispanic or Latino</td>
<td>109</td>
<td>41.0%</td>
</tr>
<tr>
<td>Native American or American Indian</td>
<td>11</td>
<td>4.1%</td>
</tr>
<tr>
<td>White</td>
<td>107</td>
<td>40.2%</td>
</tr>
<tr>
<td>Multiple</td>
<td>6</td>
<td>2.3%</td>
</tr>
<tr>
<td>Persian</td>
<td>1</td>
<td>0.4%</td>
</tr>
<tr>
<td>Arab</td>
<td>1</td>
<td>0.4%</td>
</tr>
</tbody>
</table>

Materials

The Personality Battery

The personality battery was administered first, as I did not want performance on the task to negatively influence participants’ answers. Following the personality battery, participants were administered a 29-question logical reasoning and 25-
question English grammar task. The order of the logical reasoning and English grammar task were randomized to help control for order effects.

Participants completed a total of six cognitive and personality measures. Table 4 contains the list of personality and cognitive measures used in this study, the purpose of each measure, and the reliability of each measure.

**Table 4**  
*Measures of Personality and Cognition*

<table>
<thead>
<tr>
<th>Measure</th>
<th>Items</th>
<th>Purpose</th>
<th>Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overclaiming Questionnaire (Paulhus, 2005)</td>
<td>30 out of 150-items*</td>
<td>Measures overclaiming of knowledge by assessing whether participants claim knowledge of words that do not exist.</td>
<td>α = .78 (Williams et al., 2001)</td>
</tr>
<tr>
<td>NEO-PI-3 (John &amp; Srivastava, 1999)</td>
<td>44-items</td>
<td>Measures the Big-Five personality traits: Conscientiousness, Agreeableness, Neuroticism, Openness to Experience, and Extraversion</td>
<td>α = .88 to .92 (McCrae, Costa, &amp; Martin, 2005)</td>
</tr>
</tbody>
</table>
| Sapp & Harrod's (1993) Locus of Control Scale | 6-items     | Measures three facets of Locus of Control: Internal locus of control - the belief that one controls their fate. External locus of control: Comprised of the sum of two subscales.  
- Chance - the belief that outcomes are determined by random factors.  
- Powerful Others - the belief that one’s life is controlled by others. | α = .59 - .72 (Sapp & Harrod, 1993) |
| Reynold and Gerbasi's (1982) Social Desirability Scale, Short-Form C | 13-items | The Social Desirability scale measures: Self-deception - the degree to which an individual believes in their overly positive self-reports. Impression management - the degree to which people intentionally provide a more positive image to others (Paulhus, 1984). | α = .76 (Reynold & Gerbasi, 1982) |
Schraw and Dennison’s (1994) Metacognitive Awareness Inventory (MAI) 52-items  The MAI measures: **Metacognitive Knowledge**—individuals understanding of how they learn, problem solving strategies, and how they think in general.  

**Metacognitive Regulation** – an individual’s ability to understand when they are off course and how to self-correct.  

Metacognitive ability in this study is the sum of participants’ scores on the metacognitive knowledge and regulation dimensions.  

Leykin and DeRubeis’s (2010) Decision Style Questionnaire 31-items  A measure of **intuitive decision style**.  

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

α = .90 (Schraw & Dennison, 1994)

α = .72 - .91 (Leykin & DeRubeis, 2010)

**The Performance Tasks**

Two tasks were used in this study as performance measures: a 25-question English grammar questionnaire and a 29-question logical reasoning task. I selected these tasks for two reasons. First, logical reasoning and English grammar tasks have been used in previous Dunning-Kruger research (Kruger & Dunning, 1999; Pennycook et al., 2017). Second, to explore the existence of a Dunning-Kruger trait, I needed two tasks that would measure different skill sets in the participants. Therefore, I chose two tasks that had a proven history within the Dunning-Kruger field and were not too

---

*a Using only a subset of 30-items from the Overclaiming Questionnaire is supported by prior research (Williams et al., 2001).

*b These dimensions are specified by Harrison and Vallin (2016).

---

4 The names and genders of individuals, locations, and dates were changed in the logic and English grammar questions to prevent participants from easily locating the question’s answers online.
strongly correlated with one another. Participant’s ability on the logic and grammar tasks (e.g., task scores) had only a medium correlation, $r(266) = .42, p < .001.$

**English Grammar (Grammar Task).** The questions for the English grammar task were derived from McGraw-Hill’s 2nd ACT Practice Tests (Dulan, 2008). The questions covered punctuation and comma use, proper tense, adding to or clarifying paragraphs or sentences, sentence structure, and word choice. ACT and SAT grammar questions of similar caliber were used in the original Dunning-Kruger work (Kruger & Dunning, 1999). One question from the English grammar task, question 19, was dropped from analyses due to a technical error.

**Logical Reasoning (Logic Task).** As the goal of the logical reasoning task was to assess logical reasoning abilities, the logical reasoning task was comprised of several well-established logical reasoning batteries. These batteries were combined to ensure enough questions were available to power our item-by-item analyses.

The logical reasoning task included a combined 11 questions from the Frederick (2005), Thomson and Oppenheimer (2016), and Toplak, West, and Stanovich (2014) Cognitive Reflection Tests (CRT), 15 questions from the Heuristics & Biases battery (Toplak, 2011), and three additional probability questions from Tversky and Kahneman (1974).

The overall theme of the logic battery was to test participants’ logical reasoning abilities. Therefore, I deliberately sampled a wide variety of challenging problems that test for different types of well-known reasoning errors (e.g., ignoring base rates, understanding probabilities of events, and logical fallacies). In addition, both
Frederick’s version of the CRT and Toplak’s Heuristics & Biases battery had been used in Dunning-Kruger research (Pennycook et al., 2017).

**Types of Performance Misestimation**

Throughout the rest of this thesis, I will be referring to the misestimation of the number questions answered correctly at the end of the task as *Global Overestimation* (Ehrlinger et al., 2008). Overstating one’s percentile ranking will be referred to as *Overplacing* (Moore & Dev, 2008). Item-by-item estimates that exceed accuracy levels will be referred to as *Overprecision* (Moore & Dev, 2008).

For both the Logic and Grammar tasks, participants were asked to give pre- and post-assessments of their performance, as well as provide the item-by-item accuracy estimates. The pre- and post-estimates asked participants to 1) estimate how many questions they would/did answer correctly (global accuracy estimation), 2) to estimate out of 100 other students completing the task, how many students they believed they would/did score higher than (overplacing), and 3) how difficult they felt the task would be/was (task difficulty).

**Calculating Overall Scores and Outcome Measures for Plotting and Analysis**

Rather than analyzing the data separately for the logic and grammar tasks, the majority of our analyses will focus on participants’ “overall scores” and patterns of misestimation. Overall task scores and participants’ overall performance estimates were calculated by summing participants’ scores and estimates for the logic and grammar tasks. Table 5 shows how the estimations and scores were aggregated for each of our outcome measures for plotting and analysis purposes. By focusing on
overall patterns of misestimation, I can determine which traits are predictive of overconfidence while mitigating some of the logic and grammar task-specific effects. Using overall scores also gave the statistical analyses more power. In addition, per Kruger and Dunning (1999) convention, participants were divided into four quartiles based on their overall performance. This division was achieved using the “percentile_rank” function in R statistical software.

**Table 5**

*How Estimates and Accuracy were Aggregated for Plotting and Analysis*

<table>
<thead>
<tr>
<th>Measure</th>
<th>Estimation Calculation</th>
<th>Accuracy Calculation</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global Estimation</td>
<td>Σ Estimated Task Scores</td>
<td>Σ Task Scores</td>
<td>Equation derived from Wang et al., (2012)</td>
</tr>
<tr>
<td>Overplacing</td>
<td>Σ Est. Task Percentile Ranking / Number of Tasks</td>
<td>Achieved Percentile = Percentile ranking (Σ Task Scores)</td>
<td>Equation derived from Kruger &amp; Dunning (1999)</td>
</tr>
<tr>
<td>Overprecision</td>
<td>Σ Confidence Judgments / Total Number of Task Questions</td>
<td>Σ Task Scores / Number of Total Task Questions</td>
<td>Equation derived from Wang et al. (2012)</td>
</tr>
</tbody>
</table>

After calculating participants’ overall scores and determining which quartile each participant belonged to, I then calculated the dependent measures (e.g., Global Estimation scores, the item-by-item Overprecision, and the percentile-based Overplacing) at the overall level. I did this by taking participants’ aggregated estimates and subtracting their aggregated accuracy across the logic and grammar questionnaires. Table 5 shows the formulas used to calculate the outcome measures. Table 6 contains the mean and standard deviations for the three outcome measures for each quartile, as well as the score range for each quartile.
Table 6

Mean and Standard Deviations for Outcome Measures by Performance Quartile

<table>
<thead>
<tr>
<th>Overall Quartile Range</th>
<th>N</th>
<th>Mean Overall Task Score</th>
<th>Mean Overall Global Estimation*</th>
<th>Mean Overall Overplacing*</th>
<th>Mean Overall Overprecision*</th>
</tr>
</thead>
<tbody>
<tr>
<td>13-24 (Q1)</td>
<td>71</td>
<td>21.25 (2.50)</td>
<td>3.92 (10.20)</td>
<td>0.30 (0.24)</td>
<td>0.26 (0.17)</td>
</tr>
<tr>
<td>25-29 (Q2)</td>
<td>69</td>
<td>27.06 (1.45)</td>
<td>2.97 (8.12)</td>
<td>0.11 (0.18)</td>
<td>0.22 (0.14)</td>
</tr>
<tr>
<td>30-34 (Q3)</td>
<td>66</td>
<td>31.76 (1.41)</td>
<td>-0.24 (9.16)</td>
<td>-0.11 (0.20)</td>
<td>0.12 (0.15)</td>
</tr>
<tr>
<td>35-48 (Q4)</td>
<td>62</td>
<td>38.03 (2.82)</td>
<td>-1.23 (6.57)</td>
<td>-0.26 (0.19)</td>
<td>0.06 (0.13)</td>
</tr>
</tbody>
</table>

* Positive values indicate overconfidence; negative values indicate underconfidence.
IV. Results

All data analyses were conducted with R, the R-based user interface Jamovi, and Excel. This section starts with a description of the over- and underestimation patterns for each outcome measure. Following this, I will address my research aims which were to 1) understand who is most likely to over or underestimate their performance and 2) to understand the source of this misestimation using hierarchical regression analyses.

Participants’ Ability to Accurately Assess Their Performance

To compare participants’ estimated performance to their actual performance, I constructed a series of “Dunning-Kruger plots,” following Kruger and Dunning (1999). Dunning-Kruger plots are constructed by plotting each quartile’s mean overall estimated performance against their overall mean achieved performance. In keeping with Kruger and Dunning’s (1999) tradition, I will be analyzing the difference between the top and bottom quartile performers, as they are theorized to show the greatest difference in performance misestimation.

As shown in Figure 5, participants’ percentile ranking estimates compared to their achieved percentile matches the typical Dunning-Kruger pattern. Participants in the bottom quartile overestimated their overall percentile ranking ($M_{Estimated} = 43.73^{th}$) when compared to their actual ranking ($M_{Achieved} =13.39^{th}$), paired $t(70) = 10.87, p < .01$. Meanwhile, individuals in the top quartile underestimated their percentile ranking ($M_{Estimated} = 88.6^{th}$) compared to their achieved percentile ($M_{Achieved} = 62.74^{th}$), paired $t(61) = -10.47, p < .001$. We have therefore replicated the standard Dunning-Kruger effect in this experiment (e.g., Kruger & Dunning, 1999).
**Figure 5**

*Overall Overplacing*

![Graph showing estimated versus achieved percentiles across quartiles.](image)

*Note.* This plot represents percentile ranking estimates, as measured by asking participants "out of 100 other psychology students completing this task, how many do you believe you would score higher than?" prior to and after the task. Error bars represent 95\% confidence intervals.

Participants in both the bottom and top quartiles show very little difference between their pre-percentile estimates and their post-percentile estimates (Figure 5). The bottom quartile show a small, but significant, 5 percentile difference between their pre-estimates and their post-estimates, paired $t(70) = 2.55, p < .05$. The top quartile showed a non-significant 1 percentile difference, paired $t(61) = .88, p = .38$.

Figure 6 shows participants’ overall global accuracy estimates in comparison to their achieved accuracy. As expected, those in the bottom quartile still *overestimated* their performance, as they estimated they would answer 3.9 questions more than they actually did, paired $t(70) = 3.25, p < .01$. However, the top quartile non-significantly
underestimated their performance by 1.22 questions, paired $t(61) = -1.47, p = .15$. These results also replicate the original Dunning-Kruger pattern (Kruger & Dunning, 1999).

Figure 6

*Overall Global Overestimation (Out of 53 Questions)*

![Graph showing overall global overestimation](image)

Note. This plot represents the global accuracy estimates of performance, in which participants were asked to estimate how many questions they had answered correctly both prior to completing the task (e.g., “Pre-Estimated Questions Correct”) and at the end of the task (e.g., “Post-Estimated Questions Correct”). These estimates were then plotted against participants’ accuracy, out of 53 total task questions. Error bars represent 95% confidence intervals.

Participants in the bottom and top quartiles gave significantly lower post-global accuracy estimates than their pre-global accuracy estimates (see Figure 6). One interpretation is that the tests ended up being harder than participants initially expected. Participants in the bottom quartile reduced their post-estimates by 6.14 questions on average from their pre-estimates, paired $t(70) = 8.06, p < .001$.

Participants in the top quartile reduced their post-estimates by 2.69 questions from their pre-estimates, $t(61) = 4.11, p < .001$. 
As demonstrated in Figure 7, both the low and high scoring participants overestimated their performance on their item-by-item accuracy estimates. One interpretation is that participants report on their initial feelings of accuracy, rather than thinking about the likelihood that they are actually correct. I will return to this finding in the Discussion section to further unpack this interpretation. Participants in the bottom quartile overestimated their performance by 26%, paired \( t(70) = 13.05, p < .001 \), while participants in the top quartile overestimated their performance by 6%, paired \( t(61) = 3.72, p < .001 \).

**Figure 7**

*Overall Overprecision*

![Graph showing overall overprecision](image)

*Note.* This plot represents item-by-item estimates of accuracy across both tasks. Error bars represent one 95% confidence intervals. Prior estimates were not collected at the item-by-item level due to limits on the feasibility of collecting those estimates.

The three overconfidence measures described above – Overplacing, Global Overestimation, and Overprecision – were highly correlated. Global Overestimation and Overprecision share a large positive correlation of \( r(266) = .62, p < .001 \), which is
unsurprising as they both measure participant’s perception of their performance at the question-level. However, what was surprising was the degree to which Global Overestimation and Overprecision were correlated with Overplacing. Overplacing is a social comparison measure by design, which may produce additional types of misestimation error. However, the Global Overestimation and Overplacing measures shared a positive correlation of $r(266) = .53$, $p < .001$, while Overplacing and Overprecision share a positive correlation of $r(266) = .67$, $p < .001$. The high correlations between all three outcome measures reflect the fact these measures are capturing participant’s over and underconfidence in a similar manner.

**Overconfidence Patterns Across Domains**

To evaluate Prediction 1, that some individuals are especially prone to overestimating (or underestimating) their performance, I must address two components. The first component is whether some individuals consistently over or underestimated their performance across the two tasks. The second component is to determine whether an individual’s performance misestimation is predicted by specific individual traits.

To address the first component of Prediction 1, I ran a series of hierarchical linear regression models to examine whether performance misestimation on the logic task would predict performance misestimation on the grammar task, when controlling for task skill\(^5\). Task skill was used as a control variable, so that I could assess whether overestimation was occurring beyond what task skill could account for. Kruger and

---

\(^5\) Task skill is measured by participants’ task score (for Global Overestimation and Overprecision) and their achieved percentile (Overplacing).
Dunning’s theory stipulates that task skill is the sole predictive factor of whether or not misestimation will occur. Therefore, if skill is the sole predictive factor of overconfidence, then overconfidence should be mostly accounted for by domain ability. Overconfidence on the grammar task should not then be predicted by both grammar task skill and logic task overconfidence, especially considering performance on the logic and grammar tasks are only moderately correlated, $r(266) = .42, p < .001$. If the logic and grammar task performance were highly correlated, then this could indicate that performance on these tasks are drawing from the same underlying skill set in our sample.

**Overplacing**

First, I regressed participants’ Achieved Grammar Percentile (e.g., task score on a percentile scale) (step 1) and Logic Overplacing scores (step 2) on their Grammar Overplacing scores. The results of this analysis are highlighted in Table 7 below.

**Table 7**

*Regression Results Using Estimated Grammar Percentile as the Criterion*

<table>
<thead>
<tr>
<th>Predictor</th>
<th>$b$</th>
<th>$\beta$</th>
<th>$sr^2$</th>
<th>$r$</th>
<th>Fit</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Achieved Grammar</td>
<td>-.74**</td>
<td>-.68</td>
<td>.44</td>
<td>-.74**</td>
<td>$R^2 = $ .553**</td>
<td></td>
</tr>
<tr>
<td>Percentile</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Logic Overplacing</td>
<td>.23**</td>
<td>.23</td>
<td>.05</td>
<td>.41**</td>
<td>$R^2 = $ .605**</td>
<td>$\Delta R^2 = $ .052**</td>
</tr>
</tbody>
</table>

*Note.* N=268. All standardized regression coefficients are from the final step of the analyses. A significant $b$-weight indicates the beta-weight and semi-partial correlation are also significant. $b$ represents unstandardized regression weights. $\beta$ indicates the standardized regression weights. $sr^2$ represents the semi-partial correlation squared. $r$ represents the zero-order correlation. *indicates $p < .05$. **indicates $p < .01$.

Tests to ascertain whether the data met the assumption of collinearity indicated that multicollinearity was not a concern in this analysis (Achieved Grammar Percentile, Tolerance = .94, $VIF = 1.07$; Logic Overplacing, Tolerance = .94, $VIF = 1.07$). The data also met the assumption of independence of errors (Durbin-Watson value = 2.08).
Grammar Achieved Percentile, which represents participant’s score ranking, predicted 55.3% percent of the variance in Grammar Overplacing scores. This finding essentially represents the classic Dunning-Kruger effect: the lower the skill, the more people overestimate. Critically, overplacing on the logic task significantly predicted another 5.2% of the variance in Grammar Overplacing scores. In other words, individuals who overplaced their performance on the logic task were more likely to overplace their performance on the grammar task, \( \beta = .23, p < .01 \).

**Global Overestimation**

For the second analysis, I regressed participants’ Grammar Task Score (step 1) and their Logic Global Overestimation scores (step 2) on their Grammar Global Overestimation scores. The results of this analysis are displayed in Table 8.

### Table 8

**Regression Results Using Grammar Global Overestimation as the Criterion**

<table>
<thead>
<tr>
<th>Predictor</th>
<th>( b )</th>
<th>( \beta )</th>
<th>( sr^2 )</th>
<th>( r )</th>
<th>Fit</th>
<th>Difference</th>
<th>( \Delta R^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grammar Task Score</td>
<td>-.40**</td>
<td>-.27</td>
<td>.07</td>
<td>-.29**</td>
<td>( R^2 = .082** )</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Logic Global Overestimation | .32**  | .41       | .16     | .42**| \( R^2 = .246** \) \( \Delta R^2 = .165** \)

Note. \( N=268 \). All standardized regression coefficients are from the final step of the analyses. A significant \( b \)-weight indicates the beta-weight and semi-partial correlation are also significant. \( b \) represents unstandardized regression weights. \( \beta \) indicates the standardized regression weights. \( sr^2 \) represents the semi-partial correlation squared. \( r \) represents the zero-order correlation. * indicates \( p < .05 \). ** indicates \( p < .01 \).

Tests to check whether the data met the assumption of collinearity indicated that multicollinearity was also not a concern in this analysis (Grammar Task Score, Tolerance = 1, \( VIF = 1.00 \); Logic Global Overestimation, Tolerance = 1, \( VIF = 1.00 \)). The data also met the assumption of independence of errors (Durbin-Watson value = 1.98).

Participants’ Grammar Task Score significantly predicted 8.2% percent of the variance in Grammar Global Overestimation scores. Overestimation on the logic task...
predicted another 16.5% of the variance in Grammar Global Overestimation scores.

Individuals who overestimated their performance on the logic task were also more likely to overestimate their performance on the grammar task, $\beta = .41$, $p < .01$.

**Overprecision**

Finally, I regressed participants’ Grammar Task Score (step 1) and their Logic Overprecision scores (step 2) on their Grammar Overprecision scores. The results of this analysis are shown in Table 9.

**Table 9**

*Regression Results Using Grammar Overprecision as the Criterion*

<table>
<thead>
<tr>
<th>Predictor</th>
<th>$b$</th>
<th>$\beta$</th>
<th>$sr^2$</th>
<th>$r$</th>
<th>$R^2$</th>
<th>$\Delta R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grammar Task Score</td>
<td>-.51**</td>
<td>-.39</td>
<td>.15</td>
<td>-.49**</td>
<td>.235**</td>
<td></td>
</tr>
<tr>
<td>Logic Overprecision</td>
<td>.49**</td>
<td>.54</td>
<td>.28</td>
<td>.61**</td>
<td>.520**</td>
<td>.285**</td>
</tr>
</tbody>
</table>

*Note. N=268. All standardized regression coefficients are from the final step of the analyses. A significant $b$-weight indicates the beta-weight and semi-partial correlation are also significant. $b$ represents unstandardized regression weights. $\beta$ indicates the standardized regression weights. $sr^2$ represents the semi-partial correlation squared. $r$ represents the zero-order correlation. *indicates $p < .05$. ** indicates $p < .01$.*

Multicollinearity was also not a concern in this analysis (Grammar Task Score, Tolerance = .97, $VIF = 1.03$; Logic Overprecision, Tolerance = .97, $VIF = 1.03$). The data also met the assumption of independence of errors (Durbin-Watson value = 2.20).

Participants’ Grammar Task Score significantly predicted 23.5%, percent of the variance in Grammar Overprecision scores. Overprecision on the logic task predicted another 28.5% of the variance in Grammar Overprecision scores. Participants who were overprecise in their estimates on the logic task were more likely to be overprecise in their performance on the grammar task, $\beta = .54$, $p < .01$. 
In summary, all three overconfidence measures showed that overconfidence on one task was a valuable predictor of overconfidence in the other task, even after taking into account task skill.

**Testing the Hypothesized Theoretical Pathways of Overconfidence**

As a first step, I assessed the fit of our foundational theories (e.g., General Metacognitive Ability and Social Desirability) in predicting overall Global Estimation and Overplacing. Because the item-by-item estimations and the Global Estimation outcome measures both capture participants’ estimates of the number of questions they had answered correctly, I did not conduct a separate regression analysis for the item-by-item estimates. Means, standard deviations, and correlations between variables can be found in Table 10. I will start with the fit of the theories to the Global Estimation outcome measure, followed by the Overplacing outcome measure.
Table 10
Means, Standard Deviations and Correlations between DV and IVs Used in Linear Regression Analyses

<table>
<thead>
<tr>
<th>Variable</th>
<th>M</th>
<th>SD</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Overall Overestimation</td>
<td>1.46</td>
<td>8.87</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Overall Overplacing</td>
<td>0.02</td>
<td>0.3</td>
<td>.54**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Overall Score</td>
<td>29.22</td>
<td>6.5</td>
<td>-.26**</td>
<td>-.75**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Metacognition</td>
<td>14.44</td>
<td>3.4</td>
<td>-.04</td>
<td>0.01</td>
<td>-.16**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Social Desirability</td>
<td>37.92</td>
<td>6.52</td>
<td>.12*</td>
<td>0.09</td>
<td>-.03</td>
<td>.17**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Intuitive Decision Style</td>
<td>17.77</td>
<td>2.87</td>
<td>.18**</td>
<td>.23***</td>
<td>-.26**</td>
<td>0.03</td>
<td>-.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Openness</td>
<td>36.26</td>
<td>5.06</td>
<td>.07</td>
<td>0.02</td>
<td>0.12</td>
<td>0.08</td>
<td>0.06</td>
<td>.17**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Neuroticism</td>
<td>25.36</td>
<td>5.67</td>
<td>-.15*</td>
<td>-.03</td>
<td>-.09</td>
<td>.28**</td>
<td>-.44**</td>
<td>-.07</td>
<td>0.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Conscientiousness</td>
<td>32.79</td>
<td>5.32</td>
<td>.13*</td>
<td>0.07</td>
<td>-.01</td>
<td>-.31**</td>
<td>.47**</td>
<td>-.04</td>
<td>0.04</td>
<td>-.38**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Agreeableness</td>
<td>33.64</td>
<td>4.77</td>
<td>.15*</td>
<td>0.10</td>
<td>-.13*</td>
<td>-.07</td>
<td>.57**</td>
<td>0.10</td>
<td>0.07</td>
<td>-.27**</td>
<td>.30**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Extraversion</td>
<td>25.78</td>
<td>5.98</td>
<td>0.04</td>
<td>0.04</td>
<td>0.01</td>
<td>-.12</td>
<td>0.04</td>
<td>0.06</td>
<td>0.12</td>
<td>-.22**</td>
<td>.15*</td>
<td>0.11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. External L.O.C.</td>
<td>39.68</td>
<td>7.26</td>
<td>.17**</td>
<td>.22**</td>
<td>-.12*</td>
<td>-.12*</td>
<td>.29**</td>
<td>.14*</td>
<td>.16**</td>
<td>-.25**</td>
<td>.39**</td>
<td>.14*</td>
<td>.21**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. Internal L.O.C.</td>
<td>11.52</td>
<td>1.84</td>
<td>.13*</td>
<td>0.10</td>
<td>0.01</td>
<td>-.30**</td>
<td>0.08</td>
<td>.20**</td>
<td>-.02</td>
<td>-.01</td>
<td>.17**</td>
<td>0.11</td>
<td>0.09</td>
<td>.20**</td>
<td></td>
</tr>
<tr>
<td>14. Overclaiming Bias</td>
<td>4.24</td>
<td>1.18</td>
<td>0.09</td>
<td>0.08</td>
<td>0.04</td>
<td>-0.06</td>
<td>.14*</td>
<td>0.04</td>
<td>.20**</td>
<td>-.21**</td>
<td>.14*</td>
<td>0.02</td>
<td>0.02</td>
<td>.25**</td>
<td>0.06</td>
</tr>
</tbody>
</table>

Note. N=268. M and SD are used to represent mean and standard deviation, respectively. * indicates p < .05. ** indicates p < .01.
Global Estimation and the Fit of the Foundational Theories

To assess the fit of the foundational theories to Global Estimation patterns seen in this dataset, two hierarchical regression were conducted to examine the fit of the hypothesized Metacognition and Social Desirability pathways in predicting Global Estimation scores. Each regression analysis used mean-centered variables, as mean-centering is known to reduce multicollinearity in regression analyses with a large number of variables (Iacobucci et al., 2016). The first model was designed to assess the General Metacognitive Ability pathway's fit to Global Estimation and is shown in Table 11.

Table 11
Results of Hierarchical Regression Analysis for the Metacognitive Pathway and Global Estimation

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Block 1</th>
<th>Block 2</th>
<th>Block 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall Task Score</td>
<td>-.26***</td>
<td>-.24***</td>
<td>-.24***</td>
</tr>
<tr>
<td>Metacognition</td>
<td>.14*</td>
<td>.06</td>
<td></td>
</tr>
<tr>
<td>Intuitive Decision Style</td>
<td></td>
<td>.09</td>
<td></td>
</tr>
<tr>
<td>L.O.C. External</td>
<td></td>
<td>-.07</td>
<td></td>
</tr>
<tr>
<td>Conscientiousness</td>
<td></td>
<td>.08</td>
<td></td>
</tr>
<tr>
<td>Openness</td>
<td></td>
<td>.06</td>
<td></td>
</tr>
<tr>
<td>Overclaiming Bias</td>
<td></td>
<td>.06</td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>.26</td>
<td>.29</td>
<td>.34</td>
</tr>
<tr>
<td>R²</td>
<td>.07</td>
<td>.09</td>
<td>.11</td>
</tr>
<tr>
<td>ΔR²</td>
<td></td>
<td>.02*</td>
<td>.03</td>
</tr>
<tr>
<td>AIC</td>
<td>1917</td>
<td>1913</td>
<td>1915</td>
</tr>
</tbody>
</table>

Note. N = 268; †p < .1; *p < .05, **p < .01, ***p< .001

First, my independent variables were assessed to ensure that they were not combinations of other independent variables. Next, I assessed the variables for problems with multicollinearity. The Variable Inflation Factor (VIF) and tolerance scores were within the acceptable limits (VIF < 5, Tolerance > .69), indicating that multicollinearity was not an issue in this dataset. The dataset was
also assessed for extreme outliers (e.g., scores that were +/- 3 standard deviations away). No outliers met the standard deviation criteria and thus all data were included in the model.

In the first block of the model, Overall Task Score was entered as a control variable. As the predominant casual theory in the field of Dunning-Kruger research is that of skill, it was important to assess the fit of the General Metacognitive Ability with skill controlled for. Overall task score accounted for 7% of the variance in overall Global Estimation scores, \(F(1, 266) = 19.24, p<.001\). Fitting with standard Dunning-Kruger findings, low scorers on the task were more likely to overestimate their performance at the Global Estimation level, \(\beta = -.26, p < .001\).

In the second block of the hierarchical model, the core construct of General Metacognitive Ability from the Metacognitive Awareness Inventory was entered. General Metacognitive ability accounted for an additional 2% of the variance in Global Estimation scores, \(F(2, 265) = 12.58, p<.001\). Results indicate that individuals who rated themselves as having higher metacognitive ability were more likely to overestimate their performance at the Global Estimation level, \(\beta = .14, p = .033\). This is in contrast with the hypothesized relationship, which stated that individuals with high metacognitive ability would be less likely to overestimate their performance, not more.

In the third step of the hierarchical regression, the hypothesized individual differences were entered. These individual differences include Intuitive Decision Style, External Locus of Control (L.O.C.), the Big-Five traits of Conscientiousness and Openness to Experience (i.e., Openness), and Overclaiming Bias (e.g., overclaiming one’s
knowledge). None of the hypothesized traits were found to be significant at the p < .05 level.

The next hierarchical regression analysis examined the Social Desirability pathway in predicting overall Global Estimation scores and is highlighted in Table 12. All variables had VIF and Tolerance scores within the acceptable limits (VIF < 5, Tolerance > .65).

**Table 12**

Results of Hierarchical Regression Analysis for the Social Desirability Pathway and Global Estimation

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Block 1</th>
<th>Block 2</th>
<th>Block 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall Task Score</td>
<td>-.36***</td>
<td>-.35***</td>
<td>-.24***</td>
</tr>
<tr>
<td>Social Desirability</td>
<td>.11†</td>
<td>.05</td>
<td></td>
</tr>
<tr>
<td>Agreeableness</td>
<td></td>
<td>.08</td>
<td></td>
</tr>
<tr>
<td>Extraversion</td>
<td></td>
<td>.03</td>
<td></td>
</tr>
<tr>
<td>Overclaiming Bias</td>
<td></td>
<td>.09</td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>.26</td>
<td>.28</td>
<td>.31</td>
</tr>
<tr>
<td>R²</td>
<td>.07</td>
<td>.08</td>
<td>.09</td>
</tr>
<tr>
<td>ΔR²</td>
<td></td>
<td>.01†</td>
<td>.01</td>
</tr>
<tr>
<td>AIC</td>
<td>1917</td>
<td>1915</td>
<td>1917</td>
</tr>
</tbody>
</table>

*a all variables have been mean-centered prior to analyses.

Note. N = 268; †p < .1; *p < .05, **p < .01, ***p < .001

This analysis also controlled for skill, which was entered in the first block. Skill accounted for 7% of the variance in overall Global Estimation scores, F(1, 266) = 19.24, p < .001. Again, low scorers on the task were more likely to overestimate their performance at the Global Estimation level, β = -.36, p < .001

Social Desirability was entered into the second block of the hierarchical regression. The addition of Social Desirability into the model did not significantly predict additional variance in Global Estimation scores, ΔR² = .01, p = .58.

The hypothesized additional individual traits thought to be involved in the Social Desirability pathway were entered into the third block of the hierarchical regression.
model. These individual traits included overclaiming one's knowledge and Agreeableness and Extraversion from the Big-5 personality inventory. None of the additional hypothesized traits predicted Global Estimation scores at the p < .05 level.

The Best Model to Explain Global Estimation. Akaike Information Criterion (AIC) values were used to select which model best explains Global Estimation. AIC values quantifies the level of prediction error in a model. Therefore, a lower AIC value is preferred, as it indicates that the model in question predicts the most amount of variance with the least amount of error (Wagenmakers & Farrell, 2004). Beyond having a smallest AIC value, AIC values with a difference score of 2 or greater between models represents a greater degree of fit (Wagenmakers & Farrell, 2004).

Of the three blocks of metacognitive ability hierarchical regression, the best fitting model is skill and general Metacognitive ability (e.g., step 2 in Table 11), which had an AIC of 1913. The model with skill alone had an AIC of 1917 and the model with skill, general metacognitive ability, and the individual traits had an AIC value of 1915.

Of the three blocks of Social Desirability hierarchical regression, the best fitting model is skill and Social Desirability (e.g., step 2 in Table 12), which had an AIC of 1915. Both the models with 1) skill alone and 2) skill, Social Desirability, and the hypothesized individual traits had an AIC value of 1917.

Overall, the best model to explain Global Overestimation scores in this sample is the model containing skill and Metacognitive Ability. It has an AIC value of 1913, which is 2 points lower than the best fitting Social Desirability model of skill and Social Desirability, which had an AIC value of 1915.
**Overplacing and the Fit of the Foundational Theories**

To assess the fit of the foundational theories in predicting Overplacing patterns, I followed the same hierarchical regression procedure used in Global Estimation analyses. The first hierarchical model was designed to assess the General Metacognitive Ability pathway's fit to Global Estimation and is shown in Table 13. All variables had VIF and Tolerance scores within the acceptable limits (VIF < 5, Tolerance > .76).

**Table 13**

Results of Hierarchical Regression Analysis for the Metacognitive Pathway and Overall Overplacing

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Block 1</th>
<th>Block 2</th>
<th>Block 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall Task Score</td>
<td>-.75***</td>
<td>-.73***</td>
<td>-.76***</td>
</tr>
<tr>
<td>Metacognition</td>
<td>.13**</td>
<td>.10*</td>
<td></td>
</tr>
<tr>
<td>Intuitive Decision Style</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L.O.C. External</td>
<td>-.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conscientiousness</td>
<td></td>
<td>-.02</td>
<td></td>
</tr>
<tr>
<td>Openness</td>
<td>.09*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overclaiming Bias</td>
<td></td>
<td>.06</td>
<td></td>
</tr>
</tbody>
</table>

| R     | .75  | .76  | .77  |
| R²    | .56  | .58  | .60  |
| ΔR²   |      | .02* | .02† |
| AIC   | -107 | -115 | -117 |

a all variables have been mean-centered prior to analyses

*Note. N = 268; †p < .1; *p < .05, **p < .01, ***p < .001*

In the first block of the model, Overall Task Score was entered as a control variable. Overall task score accounted for 55.7% of the variance in overall Overplacing scores, \( F(1, 266) = 334.7, \ p < .001 \). Low scorers were more likely to overplace their performance, \( \beta = -.75, \ p < .001 \).

In the second block, the core construct of General Metacognitive Ability was entered. General Metacognitive ability accounted for an additional 2% of the variance in Global Estimation scores, \( F(2, 265) = 169.8, \ p < .001 \). Results indicate that individuals
who rated themselves as having higher metacognitive ability were more likely to 
overestimate their performance at the Global Estimation level, $\beta = .13, p = .001$.

Two additional traits, External Locus of Control orientation and Openness to 
Experience were found to predict overall Overplacing. Individuals who rated 
themselves as having an External Locus of Control orientation were less likely to 
overplace their performance, $\beta = -.11, p = .010$ while individuals who were high in 
Openness to Experience were more likely to overplace their performance, $\beta = .09, p < 
.033$.

The final hierarchical regression analysis examined the Social Desirability 
pathway in predicting overall Overplacing scores and is highlighted in Table 14. All 
variables had VIF and Tolerance scores within the acceptable limits (VIF < 5, Tolerance 
> .65).

### Table 14

Results of Hierarchical Regression Analysis for the Social Desirability Pathway and Global 
Estimation

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Block 1</th>
<th>Block 2</th>
<th>Block 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall Task Score</td>
<td>-.75***</td>
<td>-.73***</td>
<td>-.75***</td>
</tr>
<tr>
<td>Social Desirability</td>
<td>.07†</td>
<td>.08</td>
<td></td>
</tr>
<tr>
<td>Agreeableness</td>
<td></td>
<td>-.05</td>
<td></td>
</tr>
<tr>
<td>Extraversion</td>
<td></td>
<td>.05</td>
<td></td>
</tr>
<tr>
<td>Overclaiming Bias</td>
<td></td>
<td></td>
<td>.09*</td>
</tr>
<tr>
<td>R</td>
<td>.75</td>
<td>.75</td>
<td>.76</td>
</tr>
<tr>
<td>$R^2$</td>
<td>.56</td>
<td>.56</td>
<td>.57</td>
</tr>
<tr>
<td>$\Delta R^2$</td>
<td></td>
<td>.00</td>
<td>.01†</td>
</tr>
<tr>
<td>AIC</td>
<td>-107</td>
<td>-107</td>
<td>-109</td>
</tr>
</tbody>
</table>

*all variables have been mean-centered prior to analyses

Note. $N = 268$; †$p < .1$; *$p < .05$, **$p < .01$, ***$p < .001$

In the first block of the model, Overall Task Score was entered as a control 
variable. Overall task score accounted for 55.7% of the variance in overall Overplacing
scores, $F(1, 266) = 334.7, \ p < .001$. Low scorers were more likely to overplace their performance, $\beta = -.75, \ p < .001$.

Social Desirability was entered into the second block of the hierarchical regression. The addition of Social Desirability into the model did not significantly predict additional variance in Global Estimation scores, $\Delta R^2 = .01, \ p = .057$.

The hypothesized additional individual traits thought to be involved in the Social Desirability pathway were entered into the third block of the hierarchical regression model. One hypothesized individual trait, Overclaiming Bias, was found to significantly predict overall Overplacing. Individuals who were high in Overclaiming bias, meaning they claimed knowledge of impossible people, places, and things, were more likely to overplace their performance, $\beta = .09, \ p < .024$.

**The Best Model to Explain Overplacing.** Akaike Information Criterion (AIC) values was again used to select which model best explains overall Overplacing patterns. AIC values quantifies the level of prediction error in a model.

Of the three blocks of Metacognitive ability hierarchical regression, the best fitting model is skill, general Metacognitive ability, and individual difference traits (e.g., step 3 in Table 13), which had an AIC of -117. The model with skill alone had an AIC of -107 and the model with skill and general metacognitive ability had an AIC value of -115.

Of the three blocks of Social Desirability hierarchical regression, the best fitting model is skill, Social Desirability, and the hypothesized individual differences (e.g., step 3 in Table 14), which had an AIC -109. Both the models with 1) skill alone and 2) skill, and Social Desirability had an AIC value of -107.
Overall, the best model to explain overall Overplacing scores in this sample is the model containing skill, Metacognitive Ability, and individual differences. It has an AIC value of -117, which is 8 points lower than the best fitting Social Desirability model of skill, Social Desirability, and individual differences which had an AIC value of -109.

**Exploratory Analysis**

The theorized pathways of metacognition and social desirability overlay an already assumed structure of what should and should not predict performance misestimation. However, I felt it prudent to conduct a final hierarchical regression model that contained all of the traits collected in this study unseparated by theory. This serves two functions: 1) predictors that were previously separated by theory could be assessed together and 2) additional non-hypothesized traits could be assessed for their role in misestimation. This exploratory regression includes two additional variables: Internal Locus of Control orientation and Neuroticism from the Big-5 Personality Inventory.

As a first step in the hierarchical regression analysis, I assessed the fit of our dataset to the underlying assumptions of a linear regression model. No variables were extremely correlated (e.g., close to a correlation of 1). In addition, all variables had VIF and Tolerance scores within the acceptable limits. The dataset was also assessed for extreme outliers (e.g., scores that were +/- 3 standard deviations away). No outliers met the standard deviation criteria and thus all data were included in the model.

A two-stage hierarchical multiple regression was conducted using Global Overestimation scores as the dependent variable, controlling for the effects of task score (e.g., skill). The second block was used to assess the ability of our individual
difference measures to predict misestimation. The purpose of this analysis was largely exploratory, as it is examining the effect of all variables collected prior to assessing the fit of our theoretical models. The predictors include general metacognitive ability (i.e., Metacognition), Social Desirability, Intuitive Decision-Making Style, the-Big-Five personality traits (i.e., Openness to Experience, Conscientiousness, Extraversion, Agreeableness, and Neuroticism), and Overclaiming of Knowledge. The results are summarized in Table 15.

**Table 15**

Results of Hierarchical Regression Analysis for Overall Global Estimation

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Block 1</th>
<th>Block 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall Task Score</td>
<td>-.26***</td>
<td>-.35***</td>
</tr>
<tr>
<td>Metacognition</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>Social Desirability</td>
<td>-0.02</td>
<td></td>
</tr>
<tr>
<td>Intuitive Decision Style</td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td>Openness</td>
<td>0.08</td>
<td></td>
</tr>
<tr>
<td>Neuroticism</td>
<td>-.13†</td>
<td></td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>Agreeableness</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>Extraversion</td>
<td>-0.02</td>
<td></td>
</tr>
<tr>
<td>L.O.C. Internal</td>
<td>0.08</td>
<td></td>
</tr>
<tr>
<td>L.O.C. External</td>
<td>-0.01</td>
<td></td>
</tr>
<tr>
<td>Overclaiming Bias</td>
<td>0.04</td>
<td></td>
</tr>
</tbody>
</table>

| R         | 0.26 | 0.37 |
| R²        | 0.07 | 0.13 |
| ΔR²       | .07† |      |
| AIC       | 1917 | 1919 |

*all variables have been mean-centered prior to analyses

*Note. N = 268; †p < .1; *p < .05, **p < .01, ***p< .001

The regression equation accounted for 13.4% of the variance in Overall Global Overestimation, $F(1, 266) = 19.24$, $p<.001$. The first block, which contained the control variable Overall Task Score, accounted for 7% of the variance in Global Estimation scores. Fitting with standard Dunning-Kruger findings, those who had lower Overall Task Scores showed greater overconfidence in their Overall Global Estimation scores than those who scored higher on the two tasks, $\beta = -.26$, $p < .001$. 
Block 2 of hierarchical regression model was designed to examine the predictive role of all of the individual difference measures collected in this study. None of the predictors were found to significantly predict Global Overestimation scores. While not at the p < .05 level, Neuroticism was found to negatively predict Global Overestimation, $\beta = -.13$, $p = .07$.

A second two-stage hierarchical regression was conducted using Overall Overplacing scores as the dependent variable. The independent variables and the order in which they were entered remained the same from the Overall Global Estimation procedure. These results are summarized in Table 16.

### Table 16

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Block 1</th>
<th>Block 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall Task Score</td>
<td>-.75</td>
<td>-.75</td>
</tr>
<tr>
<td>Metacognition</td>
<td>.08</td>
<td></td>
</tr>
<tr>
<td>Social Desirability</td>
<td>.04</td>
<td></td>
</tr>
<tr>
<td>Intuitive Decision Style</td>
<td>-.01</td>
<td></td>
</tr>
<tr>
<td>Openness</td>
<td>.10*</td>
<td></td>
</tr>
<tr>
<td>Neuroticism</td>
<td>-.05</td>
<td></td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>-.04</td>
<td></td>
</tr>
<tr>
<td>Agreeableness</td>
<td>-.05</td>
<td></td>
</tr>
<tr>
<td>Extraversion</td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td>L.O.C. Internal</td>
<td>.07</td>
<td></td>
</tr>
<tr>
<td>L.O.C. External</td>
<td>-.08†</td>
<td></td>
</tr>
<tr>
<td>Overclaiming Bias</td>
<td>.05</td>
<td></td>
</tr>
</tbody>
</table>

| R          | .75 | .78 |
| R²         | .56 | .60 |
| $\Delta R^2$ | .05** | |
| AIC        | -107 | -114 |

*a all variables have been mean-centered prior to analyses

Note. $N = 268$; †$p < .1$; *$p < .05$, **$p < .01$, ***$p < .001$

The regression equation accounted for 60.3% of the variance in Overall Overplacing, $F(1, 266) = 334.7$, $p < 0.001$. The first block, which contained the control variable Overall Task Score, accounted for 55.7% of the variance in Global Overplacing.
In the second block of the model, one other trait emerged as significant predictor of overplacing one’s score: Openness to Experience from the Big-Five personality questionnaire. Those who were higher in Openness to Experience were more likely to overplace their performance, $\beta = .09$, $p< .05$. External Locus of Control Orientation, while not significant at the $p < .05$ level, was found to predict lower Overplacing, $\beta = -.08$, $p = .06$.

**The Best Model to Explain Global Estimation and Overplacing.** Akaike Information Criterion (AIC) values was again used to select which model best explains overall Global Estimation Overplacing patterns.

For Global Estimation, the best fitting theoretical model was that of skill and general metacognitive ability, which had an AIC of 1913. The exploratory analysis looking at all the predictors had an AIC of 1917. Therefore, we can conclude that the skill and overestimation model is still the best fitting model to explain Global Estimation.

For overall Overplacing, the best fitting theoretical model was the model containing skill, Metacognitive Ability, and individual differences. It had an AIC value of -117. The current exploratory regression analysis with all of the predictors has an AIC of -114. Therefore, we can conclude that the skill, Metacognitive Ability, and individual differences model is still the best predictive model.

**Discrimination and Sensitivity to Errors**

Typical Dunning-Kruger studies measure participants’ “calibration” to their own performance by looking at their percentile ranking or global accuracy estimates in comparison to their achieved performance. Calibration refers to the degree to which
participants’ estimated performance matches their actual performance. Perfect calibration would thus be equal to no difference between participants’ estimated and actual performance (McIntosh et al., 2019). By capturing participants’ item-by-item estimates of confidence, I could conduct two additional types of calibration analyses: comparing each quartile’s average confidence on correct trials and incorrect trials and a calibration curve analysis. Figure 8 illustrates participants’ average confidence on correct and incorrect trials.

**Figure 8**

*Average Confidence on Error and Correct Trials by Quartile*

![Graph showing average confidence on error and correct trials by quartile.]

*Note.* The plot above shows participants’ average confidence when they were correct versus incorrect across each quartile. The error bars represent 95% confidence intervals.

---

6 Calibration curves are created by plotting the proportion of times participants were correct for each confidence level. For example, let’s suppose a participant gave a confidence rating of 10% ten times across the course of the experiment. Three of the ten times they gave the confidence rating of 10% they were correct. Therefore, their proportion of correct for the 10% confidence level would be .3 or 30%.  

---
Interestingly, participants in all quartiles show only a small difference in confidence between when they are correct versus incorrect. Bottom quartile participants show an 8% difference between their confidence on correct trials and confidence on error trials, $t(70) = -7.79, p < .001$. The top quartile shows a 13.4% difference between their confidence on correct and error trials, $t(61) = -14.15, p < .001$. This result demonstrates that those in the bottom quartile show less confidence when they are wrong than their higher skilled peers. Therefore, regardless of skill level, individuals across the board may have a poor ability to detect when they are incorrect. Another thing to note is that participants’ show a remarkably similar level of confidence when they are incorrect. There is only a 5% difference in average confidence between those in the bottom and top quartiles on error trials.

The second calibration analysis, the calibration curve, allows us to examine for each confidence level, how often the participants in each quartile were correct. Figure 9 is a hypothetical example of a calibration curve. The black line in the figure represents the “perfect calibration line,” which would occur only if participants’ confidence judgments exactly matched their accuracy. For example, for every judgment in which participants gave 80% confidence, they should get 80% of those answers correct. If participants are overconfident, their calibration line will be similar to or below the red line, as participants’ accuracy would lower than their stated confidence. For example, if a participant gave a confidence rating of 50% and they were correct 30% of the time, they would be overconfident by 20%. In the case of underconfidence, participants’ calibration line similar to or above the blue line in Figure 9, as participants’ accuracy would be higher than their stated confidence.
Figure 10 shows the actual calibration curve for the present sample. Each quartile was underconfident in their accuracy at the lower confidence levels (e.g., 0% - 30% confident), while all quartiles showed a degree of overconfidence at the higher confidence levels (e.g., 70% - 100% confident). Participants in Quartiles 1 and 2 showed a greater degree of overconfidence at the higher confidence ratings than those in the top two quartiles. For example, when those in the top quartile gave a confidence rating of 100%, they were correct 86% of those times. However, those in the bottom quartile were correct only 51% of the times they gave a 100% confidence rating.
Included in Figure 10 at the bottom of the graph, are the number of times each quartile gave each confidence rating across the 53 logic and grammar task questions.

**Figure 10**

*Calibration Curve Showing Proportion Correct for Each Confidence Level Across Quartiles*

*Note.* The black diagonal line represents the “perfect calibration line,” while the colored dashed lines show each quartile’s proportion of correct answers for each confidence level. The solid-colored lines represent the proportion of times each confidence level was given across the 53 task questions. The proportion of the dashed lines that fall below the perfect calibration line represent overconfidence, while the proportion of the line that falls above the perfect calibration line represent underconfidence. The error bars represent 95% confidence intervals.
V. Discussion

This research supported the hypothesis that there are some individuals who are more prone to misassessing their performance. My first prediction stated that if some individuals are prone to overconfidence in their assessments, then they would do across task domains. In addition, these patterns of misestimation would be predicted by specific individual traits.

Our findings confirmed that misassessing one’s performance in one domain (e.g., logic) was predictive of misassessing one’s performance in other domains (e.g., English grammar). In addition, I found that some individual traits were predictive of performance misassessment. While it is evident that some individuals are prone to misassessing their performance, I could not fully explain these individual differences with the personality and cognitive measures used in this study.

It is possible that additional individual traits, beyond those considered in the present study, are needed to capture individual differences, particularly at the global accuracy level. Our hierarchical linear regression model accounted for only approximately 13% of the variance in Global Estimate scores. Several additional traits, such as intelligence, working memory capacity, self-esteem, additional decision styles could be assessed for their role in the Dunning-Kruger effect. Intelligence could affect individual’s metacognitive ability, as individuals with higher intelligence not only process information more rapidly than their lower intelligence peers, but they also have extra cognitive resources available to execute metacognitive monitoring activities (Conway et al., 2002; Ohtani & Hisasaka, 2018). Working memory capacity may help determine the accuracy of one’s confidence judgements, as the ability to make an
accurate confidence judgment relies on the ability to maintain and manipulate
information in real time (Kyllonen & Christal, 1990). People with high self-esteem may
also be especially overconfident because they find it easy to think positively about
themselves (Suls, Lemos, & Stewart, 2002), and self-esteem has been implicated in the
expression of the better-than-average effect (Zell et al., 2020).

In addition, the decision-making styles of satisficers (i.e., individuals who aim to
attain merely a satisfactory outcome) versus maximizers (i.e., individuals who aim to
attain the best possible outcome) may also affect individuals’ predisposition to over or
underconfidence. Jain, Bearden, and Filipowicz (2011) argued that individuals with the
tendency to maximize are especially overconfident, as some previous research has
shown that maximizers tend to believe that they are more skilled than they actually are.

Future work will examine the role of these traits in the Dunning-Kruger effect.

**Testing the Underlying Causal Theories of the Dunning-Kruger Effect**

In the following sections, I will review the relative fit of the Metacognitive ability,
Social Desirability, and Dual-Burden accounts in explaining the Dunning-Kruger effect.

*Fit of the Metacognitive Ability Account*

I had theorized that if overconfidence in one’s assessments is due to generally poor
metacognitive ability, then measures of metacognition should predict
misestimation across domains, when controlling for skill (Prediction 1.d.). Through the
metacognitive pathway, the Dunning-Kruger effect was theorized to also be predicted
by one or more of the following correlates of low metacognitive ability: 1) intuitive
decision style, 2) overclaiming one’s knowledge, 3) low of extraversion, 4) low
conscientiousness, and 5) an external locus of control orientation. I’ll first discuss our
findings regarding the core construct of Metacognition Ability, followed by a discuss of the hypothesized associated traits.

Paradoxically, our measures of Metacognitive Ability from the Metacognitive Awareness Inventory positively predicted participants’ Global misestimation and overall Overplacing. This is the opposite of the theorized relationship, which was that participants who are higher in Metacognitive ability would be less overconfident in their abilities, not more. It is possible that participants who rated themselves highly in Metacognitive Knowledge were overconfident both on the tasks and in their metacognitive abilities. It is possible that individuals who are overconfident in their abilities are also prone to misestimating their metacognitive abilities in self-report measures. An objective measure of metacognitive ability may be necessary to explore the role of Metacognitive ability in performance misestimation.

Two of the theorized individual difference traits, External Locus of Control Orientation and Openness to Experience, were also found to predict performance misestimation at the Overplacing level. Those who were high in External Locus of Control were hypothesized to have greater overconfidence in their abilities. However, the opposite result was found in this research. Those who identified themselves as having an External Locus of Control orientation were more likely to underplace their ranking in comparison to others.

On the other hand, those who were high in Openness to Experience were more likely to overplace their performance, which is consistent with findings in prior confidence research (Buratti, Allwood, & Kleitman).
**Fit of the Social Desirability Account**

The Social Desirability hypothesis, which suggests that individuals know their true performance but choose to purposefully misestimate, was also not supported in this research. Neither Social Desirability nor the correlated traits of Agreeableness and Extraversion predicted overall Overplacing or Overall Global Overestimation. Coupled with the calibration curve analysis and examining participants’ confidence on error trials and correct trials, this may suggest that participants remain *unaware* of their true performance and therefore do not purposefully inflate their performance assessments.

Of note, one individual difference, one’s propensity to overclaim their knowledge (i.e., Overclaiming Bias), was found to predict overplacement of one’s social ranking. I had originally hypothesized that individuals overclaim their knowledge due to Social Desirability, meaning that they overclaimed their knowledge to appear in a better light to others. However, due to the fact that the core construct of Social Desirability did not predict overconfidence on either outcome measure, overclaiming one’s knowledge may represent another form of mental error.

Rather than individuals claiming knowledge of a non-existent object or entity to look better, individuals who overclaim may experience a false feeling of familiarity, which leads them to claim knowledge of impossible objects, people, and locations. It is therefore possible that the tendency to overclaim may be caused by an unconscious over-reliance on feelings of familiarity, rather than overclaiming to increase positive appraisals. Overclaiming one’s knowledge may then be directly related to overconfidence because these forms of mental errors may share a similar causal root (i.e., a false feeling of familiarity).
**Fit of the Dual-Burden Account**

Kruger and Dunning’s (1999) seminal theory, that skill is the sole factor driving individuals to over or underestimate their performance, was also not fully supported. When controlling for grammar skill-level, participants’ misestimation on the logic task was predictive of performance misestimation on the grammar task across the Overplacing, Global Overestimation, and Overprecision outcome measures. This suggests that there is an underlying mechanism unrelated to skill that also leads some individuals to be overconfident and other individuals to be underconfident in their performance. Note that these large individual differences are obscured in the standard analysis procedure for the Dunning-Kruger field, which shows means for entire quartiles rather than for individuals.

Second, in the hierarchical regression analysis for Global Overestimation, skill accounted for only approximately 7% of the variance in Global over and underestimation. If skill was the driving factor, it is likely that skill would have accounted for more variance than it did. Therefore, we can conclude that skill is clearly not the only important factor in producing Global Overestimation. In addition, Metacognitive Ability as measured by the MAI also predicted participant’s misestimation, though not in the hypothesized direction.

In contrast to Global Overestimation, participant’s skill was tightly tied to whether participants over or underplaced their overall percentile ranking, as skill accounted for approximately 56% of the variance in Overplacing. While skill accounted for a large amount of variance in Overplacing scores, it was still not the sole factor. Four individual traits were also found to predict participants’ Overplacing scores:
Metacognitive Knowledge, Openness to Experience from the Big-5 Personality Inventory, Overclaiming Bias, and External Locus of Control Orientation. Therefore, skill alone was not enough to fully explain Overplacing patterns.

**Alternative Theories of the Dunning-Kruger Effect**

We believe that there is a fourth causal theory for the Dunning-Kruger effect, one that could only be tested by capturing item-by-item estimates of performance. We have termed our new causal theory of the Dunning-Kruger effect the “error blindness” account. As evident in our analysis of participants’ average confidence on correct and error trials (Figure 8), participants across all performance levels are similarly highly confident when making errors. I theorize that individuals are generally blind to mental errors. They generally do not realize when they have made a bad assumption, miscalculated, or retrieved a false memory. Critically, this could be an attribute of all humans, regardless of their skill level.

To fully unpack this theory, we must first understand when individuals are most likely to be alerted to their errors. People do sometimes receive an internal sense or “gut-feeling” that something went wrong when they are in error. This can be indexed, for example, by the error-related negativity (ERN; Gehring et al., 2012). However, ERN does not occur with *every error*, only ones that produce a conflict between what occurred in the physical world and the idea of what should have occurred (Yeung et al., 2004). For example, an ERN is likely to occur when a person accidently puts their milk in the pantry instead of the refrigerator or accidently deletes their master’s thesis word document from their computer. But an ERN would not necessarily occur for most
mental miscalculations, where all we know is the outcome of our mental processes and have nothing to compare it against.

As I discussed earlier in this paper, people can suffer from False Feelings of Knowing, in which there is no conflict between what has occurred and what should have occurred. When unsure of an answer to a problem, it is possible that individuals may unconsciously infer answers that appear reasonable or expected. Because these inferential processes are mostly unconscious, the individual remains blind to the fact that they have produced an educated guess rather than directly retrieving actual knowledge from memory. An EEG Dunning-Kruger study conducted by Muller (2019) found that individuals who overestimate their performance have a larger late parietal component when making performance estimates, which is associated with familiarity-based episodic memory processing. This may suggest that when individuals are in error, they may have been merely familiar with the answers they gave. Unfortunately, these individuals may assume that they did know the answer, because there is no error signal associated with their incorrect answer.

Therefore, the reason we see the Dunning-Kruger effect in the manner that we do (e.g., patterns of overconfidence in lower scorers and patterns of underconfidence in higher scorers) may not be due to true under/overconfidence. Rather, everyone is underconfident when they're right and overconfident when they're wrong. We may differ only in the relative proportion of these two states. Participants in the lower quartiles make a greater proportion of errors – which makes them appear more overconfident than those in higher quartiles. On the other hand, those in the top quartile are mostly correct by definition, which means that their underconfidence on
their correct trials makes them appear underconfident. This effect was not noted in prior Dunning-Kruger research, because they did not have item-by-item estimates of performance to examine confidence across error and correct trials.

The error blindness model may also help to account for why our participants had such different patterns of over and underconfidence across our three outcome measures. I theorize that overconfidence depends critically on whether individuals are asked to assess their own performance (e.g., Global accuracy and item-by-item accuracy estimates) or rank their performance in comparison to others (e.g., percentile ranking estimates). In our sample, participants were relatively well-calibrated when estimating how many questions they were likely to answer correctly at the global accuracy level (Figure 6) but poorly calibrated to how well they compared to other students completing the task (Figure 5) and overconfident when asked to assess their performance at the item-level (Figure 7).

This difference between Global and item-by-item misestimation may be due to the fact that individuals' have a better sense of how they perform overall, because they can compare their performance to past averages in the domain. In other words, at the Global level, they rely on base rates. Participants are likely to be worse at estimating at the item-level because they have less base-rate information that they can leverage to correctly assess their performance. In essence they are “blind” to their errors trial-by-trial but can strategically correct for this by taking into account base rates. For example, a low scorer might feel that they knew most of the answers, but then correctly infer that they will once again receive a low overall score.
As for percentile ranking estimates, skill is undoubtably a large factor in participant’s ability to estimate how they compare to others. However, skill alone is not the only factor that can drive this relationship. Personality factors like receptivity to novel situations and acceptance of uncertainty can also drive individuals to be too confident in how well they rank in comparison to others.

It is also possible that individuals are prone to two sources of error when estimating their percentile ranking: misestimating their own performance and that of others. As we have seen, while participants do overestimate their Global task performance, they are still more calibrated regarding their Global task performance than regarding their percentile. This may be because they have base-rate information to work from for their global accuracy estimates. That is, they more frequently receive and pay attention to feedback about their performance than about others’ performance, which they would need to infer their percentile ranking. It is possible therefore that the true source of the Overplacing error is that participants lack a useful reference point for others’ performance. This misunderstanding of others may then be driving the over and underestimation seen in most Dunning-Kruger studies, as these studies use percentile estimates as their primary performance estimate measure. In future studies, I plan to also ask participants to estimate others performance to tease apart the source of the Overplacing error.

**Research Limitations**

While participants were instructed that the numbers for the item-by-item estimates corresponded to a prediction about accuracy, it is possible that some subjects misunderstood how to correctly use the confidence ratings. However, it is important to
Note that correlations between the percentile rankings, global accuracy estimates, and item-by-item confidence estimates indicated a strong correspondence. Therefore, it is likely that the item-by-item estimates of accuracy are properly capturing participant’s over and underconfidence at the item-level.

In addition, this study has been conducted only on University of New Mexico undergraduate students. Future research should be conducted in additional populations to determine how well these findings generalize to diverse populations.

Summary

In sum, this research brought several innovations to the field of Dunning-Kruger research. First, this study was the first Dunning-Kruger study to explore individual differences in the expression of the Dunning-Kruger effect. Our results indicate that when individuals over or underestimated their performance, they showed a consistent pattern across tasks. Nevertheless, the personality and cognitive measures used in this were unable to adequately explain why some individuals are prone to misestimation. Our findings indicate that Openness to Experience was found play a role in the propensity to be overconfident, though this was limited to misestimating one’s percentile standing.

Second, this study was the first of its kind to pair pre-assessments of performance, performance assessments at the item-level, and assessments at the end of the task with knowledge-based tasks. By using additional performance estimates, I could examine participant’s understanding of their baseline performance (pre-assessments), how they felt they were doing at each stage of the tasks (item-by-item estimates), and how they felt after completing the tasks (post-assessments). As a result,
a new view of the Dunning-Kruger effect has emerged, the error blindness view of the Dunning-Kruger effect.

Ultimately, participants in this sample were much better at estimating their global accuracy performance, likely because they can use prior base rates to approximate a relatively reasonable estimate. Individuals are much more overconfident at the item-by-item-level because they do not have a good base rate in which to work from at a level that granular. Finally, overplacing one's percentile appears to operate differently than Global and item-by-item estimates, likely because there are two potential sources of error: misestimating oneself and misestimating others. This has great implications for the field of Dunning-Kruger research, as a large proportion of studies use percentile estimation as their only source of performance estimation. Future work will further explore the role of misestimating others as a source of error in the Dunning-Kruger effect.

These additional assessment measures also allowed us to examine participants’ calibration to their own performance. These calibration analyses explored participants’ average confidence on error and correct trials and their likelihood of being correct at each confidence level. These results support a new theory of the Dunning-Kruger effect: the error blindness model. This theory states that individuals lack an error signal when they are misestimating their performance, because they’re not truly recalling information. Rather, they are suffering from False Feelings of Knowing because they are using processes of familiarity to select their answer. Individuals remain “blind” to the fact that they have used familiarity processes and that they may be incorrect.
VI. References


Leykin, Y., & DeRubeis, R. J. (2010). Decision Style Questionnaire [Database record]. Retrieved from PsycTESTS. doi: [https://dx.doi.org/10.1037/t62224-000](https://dx.doi.org/10.1037/t62224-000)


https://doi.org/10.1016/j.socscimed.2018.06.032


https://doi.org/10.1016/j.intell.2004.02.001


