

Aptitude and Overconfidence via Self Assessment:  
A Survey Based Experiment on Prediction and the BAROCO Short.

John W. Lee

Southern New Hampshire University

Author Note:

John W. Lee is a graduate student in the Master of Science in I/O Psychology program.

This final project represents a research manuscript presenting an introduction to the chosen research topic, a literature review, research question, hypotheses, methods, and results.

Overconfidence as a factor of self assessment is a widely studied human phenomenon (Dunning, Heath, & Suls 2004, 2018; Freund & Kasten 2012; Logg, Haran, & Moore 2018, Zell & Krizan 2014). Self concept, and exaggerated levels of confidence have been shown to vary across cultures and produce a wide array of impacts on both present circumstances and future events, in addition to impacting outcomes, and behaviors (Dunning et al. 2004, 2018; Freund & Kasten 2012; Heine, Lehman, Markus & Kitayama 1999). Additionally, self assessment through the use of surveys is a common method used by organizations seeking data from it's employees, prospective employees, clients, and leadership. The heightened focus on self assessment makes knowledge of phenomena such as overconfidence pertinent to a better understanding of self, and the subsequent advantages or disadvantages overconfidence may pose (Freund & Kasten, 2012).

The manner in which self-assessments are conducted, the extent to which certain inferences are drawn from those assessments, and the level to which we can assess the validity of the assessment measure itself, greatly depend on the researcher's ability to account for effects such as overconfidence. In terms of the individual survey taker participating in an industrial/organizational assessment, overestimating ability can have damaging consequences not only on the individual, but also on the organization to which the individual applies its resources. Therefore, research that demonstrates the ability to modify or improve the accuracy of self-assessments stands to broaden the discussion regarding the design, validity, and applicability of the assessment tools themselves (Dunning et al. 2018).

Zelle and Krizan (2014) explored the relationship between self-assessments of ability as compared to measures of objective performance. This allowed a systematic overview of self assessments or evaluations of ability, such as intelligence tests, to be compared to performance measures such as grades, evaluations, or test scores. Combined with the findings from Dunning

et al (2004, 2018) that show a similar correlation between level of ability and accuracy of prediction, the research herein plans to add to the body of experimentally derived knowledge that explores the correlation between aptitude and overconfidence, while simultaneously manipulating the latter through benchmark priming. This will, on the one hand, provide additional research to support the findings that aptitude positively correlates to accuracy of prediction, while also providing a useable method to modify the prevalence of overconfidence in self assessment measures.

Findings that support benchmarking as a means to modify overconfidence in self-assessment lend valuable contributions to discussions regarding the explanation of overconfidence, and it's cognitive underpinnings, such as those discussed in Logg et al. (2018), who recently found that previous research may have exaggerated motivation's role in overconfidence. It is my opinion, that self assessments in the form of surveys do not accurately reflect the motivations of the respondent (as found with the study of overconfidence), flaws in self assessment, or the effect of benchmarking on the level of overconfidence displayed. This alludes to the prediction that if overconfidence were truly a result of maintaining self image, rather than a possible desire to project confidence, its impact would be less effected by benchmarks dissuading overconfidence. Extensive research would be necessary in order to achieve the necessary support to establish the aforementioned as a viable topic, however this research begins to gather relevant data to further develop this discussion.

### **Relevant Research Reviewed**

#### **Flaws in Self Assessment**

Dunning et al. (2004) outline overconfidence found in self assessment in a thorough, and systematic comparison of laboratory findings and real world examples that highlight the

individuals struggle to correctly assess themselves. Two major themes are identified, the first of which points to the finding that perceptions of skill, and actual skill only correlate with each other moderately, and sometimes they do not correlate at all. (Dunning et al. 2004) Second, people tend to rate themselves as better than they actually are, and overestimate the likelihood that they are correct (Dunning et al. 2004). The authors give both a broad overview of the research that makes up the body of knowledge regarding self assessment and overconfidence, as well as going into detail in an attempt to identify some causal factors. The laboratory studies that are discussed as laying the experimental groundwork for theories involving self assessment and overconfidence contain findings related to the propensity of individuals to assess themselves as above average, thereby producing an abhorrent challenge to mathematical law. If everyone were above average, there would be no average, yet Dunning et al. (2004) find research conducted in the lab and in the field that demonstrate a variety of ways in which more than half of people assessed systematically express overconfidence, overestimation, and un-realistic attitudes regarding their own objective ability as being above average.

The authors' note that often times the laboratory is pointed to as a source of these and similar effects. To this potential limitation, excellent efforts are made to include real world research conducted outside of the laboratory that also finds a propensity for errors in self assessment. Specifically the author's focus on three areas, health, education, and the workplace to find examples that correlate with the laboratory findings. This reflects the author's attempts to not only identify and describe overconfidence in self assessment, but to point out relevant and beneficial findings that could be of use to others at multiple levels across all three areas. One such finding is that of overconfidence expressed in judgment and prediction.

The authors' reference nine research studies conducted between 1965 and 2000 that give support to an overconfidence effect. This is described by Dunning et al. (2004) as placing overconfidence into the insightfulness of judgment, where the belief that decisions regarding the present are valid, and predictions regarding the future will be found to be correct. One of the studies referenced by Dunning et al. (2004) found that college students who expressed 100% certainty that answers they gave were correct were wrong approximately one out of five times. Taking this focus on education and overconfidence out of the laboratory, Dunning et al. (2004) provide strategies that have been found by researchers to have diminutive effects on overconfidence in self assessments, one of which is described as benchmarking. The practice of benchmarking implements a pre-assessment primer that allows the self assessor the opportunity to consider the performance of others at the same task they are about to perform. This process has been shown to provide stronger correlations between actual performance, and predicted performance (Dunning et al. 2004).

In 2018, Dunning, Heath, & Suls returned to flaws of self assessment, revisited some of the themes from their original work published in 2004, and drew new conclusions based on fresh research conducted across a wide variety of focus areas in psychology. The authors point to research done in neuroscience, cognitive, abnormal, and organizational psychology, among others, alluding to the issue of self assessment and the inherent flaws. Dunning et al. (2018) discusses the propensity for psychological research to seek out solutions, and in the case of self assessment, more accurate self appraisals. One technique to accomplish this, as discussed in both Dunning et al. (2004) and Dunning et al. (2018), was in regard to the methodical approach taken toward instruction as it relates to reported confidence in correct answers given. Essentially, the authors suggested that cramming, in the form of taking in large quantities of information required

to pass a subsequent test, leads to more rapid learning, and faster and more elevated levels of confidence. However, because this method also leads to higher levels of forgetting, the feelings of confidence are often an illusion. This can lead to inaccurate self assessments and overconfidence in predictions. Researchers identified by Dunning et al (2004; 2018) found that learning spread out over longer periods of time and including increased difficulty produced lower levels of satisfaction or confidence in students, but increased retention and learning (Dunning et al. 2018).

When considering the passing of time since exploring the topic of self assessment and the flaws found within, Dunning et al. (2018) presents research conducted in the subsequent years that explores the role of technology, and the prevalence of highly accessible data collection tools like smart phones and watches, and their role in improving the accuracy of self assessment. In addition, the authors point out research that is discussed later in this proposal that identifies self assessment, and inherent flaws like overconfidence, as being part of cultural normative behavior, with a high prevalence of the phenomenon of overconfidence in North American and Western European cultures as opposed to Eastern cultures (Dunning et al. 2018). Overall the authors again convey their desire to enhance self knowledge ability, improve the accuracy of self assessments, and continue to build upon what is a vast topic, explored throughout the ages (Dunning et al. 2018).

### **A Meta-Analysis Regarding Evaluation Validity**

Freund & Kasten, (2003) conducted a meta synthesis regarding self estimates specifically of cognitive abilities, and their relationship to psychometric test scores. Whereas Dunning et al. (2004; 2018) provided more general conclusions regarding self assessment, Freund & Kasten take a much more detailed and analytical look into one specific aspect of the related

phenomenon. Relying on a very similar group of studies, Freund & Kasten draw upon a foundation of self estimates, and their validity when cognitive ability is being assessed through psychometric testing. Some of the most novel, and relevant findings outlined by Freund & Kasten include finding a tendency for the validity of self estimates to improve if a number of formative factors are met. Two of these factors of particular relevance to this study relate to the impact of benchmarking found by Dunning et al. (2004). These two factors describe that people who have experience in self evaluation, and that people who are expecting to have their self evaluation abilities assessed, are both more likely to be accurate in their predictions (Freund & Kasten, 2003).

The researchers hypothesized that there is a significant positive correlation between self estimates regarding cognitive abilities and test scores produced from psychometric testing. This hypothesis seeks to show the reliability of the measures used by researchers and their associated studies within the group of 41 published studies assessed by Freund & Kasten, (2003) in this meta synthesis. The results included a 154 effect size, with correlation ranges from  $r = -.25$  to  $r = .85$ , mean =  $.32$ , and a standard deviation of  $.19$ . Support was found for the hypothesis that the relationship is positive between self estimation and cognitive abilities. This positive correlation was shown among 93% of the studies assessed. This equated to 22,256 participants across 42 studies. However, although the correlation was positive across a majority of the studies, supporting the hypothesis, the level of correlation between predicted self estimation and actual cognitive ability was  $r = .33$ . This is in line with prior research, that finds the validity of self estimates as indicators of cognitive ability to be low. More plainly, when considering self assessment strictly in terms of predictions as they relate to cognitive ability, it is found that the

predictions are not accurate, and they are less accurate as cognitive ability decreases (Freund & Kasten, (2003).

This is the first meta synthesis referenced in my proposed research, and provides a wealth of information regarding prior research conducted related to this topic. In addition, the meta-synthesis provides it's own findings that are also relevant to my study, as they pertain to measures used to study self assessment, and those measures validity. This article provides both a historical glimpse into the previous research on the topic, in addition to statistical analysis of those works allowing for certain inferences regarding methodology to be drawn (Freund & Kasten, (2003).

### **Bias Motivated**

Having a firm foundation in the related literature, as found through the assessment of literature reviews and meta analysis of Dunning et al. (2004; 2018) and Freund & Kasten (2012), the experimental research conducted by researchers from Harvard University and Ben-Gurion University of the Negev presents substantiated support regarding cognitive explanations of overconfidence (Logg, Haran, & Moore, 2018). The authors present a new definition of overconfidence, giving it two distinct forms; better than average beliefs or over placement, with better than average beliefs becoming visible in research when the use of vague measures is applied, making the overconfidence visible within a sample of individuals. The second form of overconfidence described by Logg, Haran & Moore (2018) defines it as over placement, overestimation, or over precision that can usually be observed when specific measures are used, making the bias of overconfidence visible at the individual level. The researchers designed their experimental study to explore the causes of overconfidence through the lens of this new, multi form approach to studying overconfidence, developing an empirical test of motivation as it



relates to overconfidence, and a means to measures its strength. Utilizing a well crafted series of six experiments, the Authors set out to test their hypothesis that if motivation lies at the root of overconfidence, then more overconfidence should be observed when people display or predict attributes and abilities that are considered important to them.

Of particular relevance to the current study, is experiment number six conducted by Logg, Haran & Moore (2018). Here the researchers used a 3 cell between subjects design that assessed how the level of motivation would correlate with better than average and over placement measures. The participants, 111 university teachers and students, completed a 10 question test of logic and math puzzles, with a varying level of motivation given at the onset of the test via a priming dialogue presented according to the level of motivation desired; low, medium, or high. To measure over placement, the authors had participants estimate their own score, and that of others. The results found only slight over placement evidence before the task, with predictions indicating performance better than others with a mean of .44, and standard deviation of 1.58, when in fact people outperformed others a mean of 0.0, and standard deviation 1.84. The authors point to prior research, in addition to these findings as evidence that the degree to which a measure contains a specific task, in addition to the difficulty of that task, relates directly to the level of overconfidence that will be displayed. This combined with the other five lab based, and methodologically similar experiments, led the researchers to conclude that motivation for a positive self view has a limited impact on overconfidence. Additionally, insight was gained as to the role of objective domains that are verifiable and their relationship with overconfidence, finding that the more objective, and the more verifiable a measure is, the less overconfidence will present.

### **Positive Self Regard**

Researchers from the University of Pennsylvania, University of British Columbia, Stanford University, and Kyoto University in Kyoto Japan took a deeper look into the motivations behind self assessment trends, such as overconfidence, as they relate to a cross cultural approach to the study of self (Heine, Lehman, Markus, & Kitayama 1999). This article draws very novel conclusions as an expansion of research touched on up until this point in the references previously discussed. By highlighting a unique finding regarding cross-cultural differences in self assessment, the authors introduce a new element into the conversation regarding the origins of what are found to be very common findings among researchers studying overconfidence in self assessment. The authors find trends in research conducted in Japan that point to a decreased level of self-enhancement or overconfidence, that is exemplified in the research of North American subjects who consistently display overconfidence. Not only does this research add valuable insight into potential confounds to be controlled for in the study of self assessment, such as cultural influences, it also points to a larger realism that self assessment may be more influenced by social processes than previously assumed (Heine et al. 1999).

The relevant findings presented by Heine et al. (1999) rely on the premise that overconfidence expressed in self assessment reflects a motivation by individuals to affirm a positive self image and maintain elevated self esteem. This is described by the authors as a reflection of a desire for independence, responsibility, success, and happiness, among other concepts, that seem to be common among North American subjects. In comparison to Japanese culture, that organizes a different set of desired concepts less focused on independence, and individualistic affirmations, self assessments lack the prevalence of overconfidence as a factor of maintenance of positive self image due to the cultural separation of desired self concepts (Heine et al. 1999).

This literature review provides another cache of extremely relevant findings regarding self assessment and in this case, positive self regard as it relates to cultural differences. It gives an overview of the research that has been conducted, as well as provides assessments of areas lacking research. It is the only article out of articles herein that explicitly explores the role of culture in self assessment. What this articles lacks in experimental data from the authors, it makes up for in the eloquent presentation of a large variety of studies containing extremely relevant, valid, and significant findings relevant to broadening the role of motivation in self assessment.

### **IQ, General Ability, and the BAROCO Short**

The final study presented here integrates a number of elements from the above works by Dunning et al. (2004, 2018), Logg, Haran & Moore (2018), as well as the experiments conducted by Heine et al.. (1999) Researchers from Japan and Germany designed an experiment to test the relationship between the BAROCO short, a syllogism assessment designed as a tool to assess various aspects of ability and intelligence, and general ability (Shikishima et al. 2011). The authors presented three hypotheses that assessed the mean score on the BAROCO short in terms of age, socioeconomic status, and genetic effect. A total of 560 families participated in the study, including 487 pairs of twins, with ages ranging from 13 to 68. The five questions included in the BAROCO short were selected due to a previous research study that found each of the five selected questions to have the highest prediction of complete IQ scores. The authors here give a detailed account of both the intricacies of the BAROCO measure, as well as the method of selection for the questions included in their study (Shikishima et al. 2011).

The results of the study find that the BAROCO syllogism measure accurately reflect general ability. The authors point to previous knowledge of the correlation between age,

socioeconomic status, and genetics as they relate to general ability. The findings herein show support for the BAROCO demonstrating the same, or similar correlations, providing evidence of the syllogism measure as a means to assess general ability and subsequent human intelligence. Although this study does not address overconfidence directly, it does give credence to an excellent measure of ability and intelligence, that fits a cross cultural application and eliminates potential confounds of a North American derived assessment that could contain implicit bias toward overconfidence as discussed previously in Heine et al. (1999). In addition this measure has extensive applications that point to the specificity to which it accomplishes its intended objective, making the work by Logg et al. (2018) that much more relevant.

### **Hypotheses**

Given the above research, the research question I propose is that when given the BAROCO short, will SNHU graduate students accurately predict their score, and is there a correlation between the level of prediction and the aptitude measured? Additionally, when given a primer as a benchmark, establishing overconfidence as a detriment, will survey respondents be less likely to project overconfidence through their predictions. The following two hypotheses will be explored:

*Hypothesis #1.* Respondents with high scores on the BAROCO short will have an accuracy of prediction greater than the accuracy of prediction of respondents with lower scores on the BAROCO short.

*Hypothesis #2.* Respondents exposed to priming that justifies overconfidence will display an accuracy of prediction less than the accuracy of prediction of respondents exposed to priming that dissuades overconfidence.

### **Method**

## **Participants**

The participants in this study were graduate students from Southern New Hampshire University enrolled in the PSY 510 and PSY 520 courses for the 19/20 Term 1, October to January, 2019. In order to prevent any perceived bias toward the results of this study, and those results relationship to gender, a sex question was intentionally left out of the survey. There are 27 sections of these two courses being offered by SNHU during that term, with a total of approximately 553 possible seats available. Based on a sample from the current PSY 510 section, the classes are likely to contain more females than males (84% female, 15% male). Undergraduate degrees vary within the current class, with not all students holding a bachelors degree in Psychology, however a significant number of students have reported having experience and/or a degree in psychology prior to enrolling in this course. No question was asked regarding degree's held, however respondents did report their undergraduate GPA. The inferred information was gathered through participation as a student in the current PSY 510 course, review of enrollment in upcoming terms, and review of student comments and names to estimate gender, in current PSY 510 course section. Twenty two people participated in the study, and using the percentages by gender from the current section of PSY 510, this would predict that 18 females, and four males took the survey.

## **Materials**

The survey was administered utilizing a computer based survey software called Qualtrics, and required that participants use a computer, or similar device to complete the survey. The survey was written in English, and required that each participant had the ability to select their answer from the choices given on the screen by pressing the appropriate key or button. Timing options were utilized within Qualtrics to analyze the amount of time taken on each question, as

well as limit the amount of time before the next question could be viewed, and limit the amount of time a respondent can spend on a question. Only electronic completion of the survey was available, and was only accessed by students enrolled in the PSY 510 and PSY 520 sections. The Survey had two different versions, version one and version two. Volunteer participants were randomly assigned to either version one or version two by the Qualtrics randomization feature. Of note here that the randomization of participants to version one or version two condition did not equally assign the same number of participants to each version. Fifteen participants were assigned to version one and seven participants were assigned to version two. Although unbalanced in terms of number of participants assigned, the assignment was performed by the software, and does appear to be random. The two versions were completely independent of one another, and participants were randomly assigned based on the order in which they volunteered. Randomization was a function of the Qualtrics software, and additional details on the specific randomization details are not available.

### **Procedures**

Students in the PSY 510 and PSY 520 courses were instructed by their professors to log on to the survey portal and view the available research projects soliciting volunteer participants. The perspective participants saw a display containing the title and student author of each one of the research project survey choices. The title of this proposed research study survey was be “Click here to find out your IQ in 5 short questions!”. Once a student clicked on the survey opened to an predetermined informed consent, and voluntary participation prompt from Qualtrics and SNHU, directing the student that their participation was completely voluntary, their responses are anonymous and confidential, and will in no way be able to be linked to them personally. They were also told that they may choose to stop taking the survey at anytime. If the

participant chose to continue with the study following providing informed consent, they proceeded to the first block of the survey. The first block contained an introduction to the survey, and differed between version one and version two. Both versions included an introduction that gave general instructions for taking the estimated IQ test (BAROCO Short), giving prediction responses, timing requirements, instructions to record each of the five answers on pen and paper in order to calculate estimated IQ score upon completion of test, and overall stated objectives of the test. Version 1 included no quote regarding overconfidence. Version two included a warning regarding the problems of being overconfident when making predictions. The quote read: “Caution! When people make predictions on their scores are usually too confident that their scores are accurate. That is why answers that seem correct may not be, more often than one might think. Hence, when estimating your score, be sure to make your prediction accurate enough to be 90% confident.” All remaining aspects of the survey were identical for each version. Following the first block, which will have unlimited time, the participant clicked next and was taken to a series of demographic questions that asked the respondent to identify their age range, how they thought they would score on the test, ethnicity, and gpa range. Time for these selections was also unlimited. After answering the final demographic question, the respondent will click next and be taken to question 1 of the survey.

The first hypothesis endeavors to identify the nature of the relationship between level of prediction and test score on the BAROCO Short (Shikishima et al., 2011). The BAROCO short consists of five questions, and each question was given on the survey as an individual timed question. Respondents had 1 minute to answer each one of the five questions from the BAROCO short. The 5 questions were presented in the same order for each participant. Each one of the five

BAROCO questions was immediately preceded with a prediction question, that was untimed. The prediction question was the same each time it is asked, and contained the following text:

“I predict that I will answer \_\_\_\_ out of \_\_\_\_ questions right on this IQ test”

- a. 5 out of 5
- b. 4 out of 5
- c. 3 out of 5
- d. 2 out of 5
- e. 1 out of 5

The order of the available answers was randomized to control for ordering effects. This totaled 10 questions, 5 questions from the BAROCO short assessment, and 1 question (stated above) asked immediately before each of the 5 BAROCO questions. After completing all 10 questions, or when choosing to exit the survey, the respondent was taken to the results block. Here a list of answers to the five questions was given, and there was a chart derived from the BAROCO short that gives a general aptitude or estimated IQ score that relates to the number of questions answered correctly. This scale displaying the related estimated IQ score to each one of the possible test outcomes (1 out of 5, 2 out of 5, etc) was derived from the research by Shikishima et al. (2011) that draws relevant comparisons to the BAROCO short and other intelligence tests. A disclaimer was given that a numerical estimated IQ score is not the specific intent of the BAROCO, and instead has been shown to have a strong correlation with level of estimated IQ. The disclaimer also included a link to the Wechsler Short IQ Test for respondents who desire to find out their IQ from a more widely used IQ test in the United States. (Adams et al., 1984)

### **Ethical Concerns**



Informed consent was established with each respondent individually prior to commencing the survey through a series of prompts initiated by the Qualtrics Software. Each respondent had the opportunity to exit the survey at anytime. Confidentiality was addressed by not recording any identifying characteristics of each respondent outside of the information provided. The respondents were not asked for their name, student ID, course section, or any other identifying characteristics, nor was the Qualtrics software able to capture any identifiable information from each respondent. There was no short answer opportunities for respondents to volunteer identifiable information. Questions regarding demographics included a “choose not to respond” answer choice. Specifically regarding Race/Ethnicity, an answer selection was provided that reads “I would rather not be asked this question.” The number of respondents selecting this answer was analyzed for inclusion in future research question development. The study, in its entirety was submitted to two professors at SNHU, who did not require changes to aspects of the research regarding ethical challenges. As this is a research project conducted as a requirement for a graduate psychology course, anticipated sample sizes were not predicted to be large enough to provide statistically significant results that would apply to a larger population. However, the participation of 22 graduate students far exceeded this prediction, and likely will provide more generalizable conclusions.

### **Data Analysis**

Utilizing SPSS the following variables were created: GPA, Version of survey (no warning or warning), Level of prediction, accuracy of prediction, and estimated IQ test results. Regarding prediction, two evaluations were made, level of prediction and accuracy of prediction. In order to determine the accuracy of prediction, each respondent’s data will be coded to determine accuracy of prediction for each respondent. This was achieved by converting the prediction selections in

each one of the prediction selections questions, “1 out of 5, 2 out of 5”, to a decimal.

Additionally, the pre-test question regarding level of prediction, ranging from moderately below average to moderately above average were coded as, .2-.8 respectively. This equated to six prediction scores for each participant, ranging from zero to one. The average was taken for each participant and recorded. This created an average prediction score that was then compared to each participant’s actual score on the IQ assessment. The actual scores ranged from zero to one as well. In order to calculate the accuracy of prediction, the actual score was subtracted from the average prediction value, producing a value ranging from -1 to 1, with zero representing a perfect prediction of the participants score. This final number is then reported as the accuracy of prediction. The other prediction variable was taken directly from the 1<sup>st</sup> question, which asked each participant how they predicted they would score on the upcoming assessment, with the choices ranging from “0 out of 5, to 5 out of 5” answered correctly. The same scale as accuracy of prediction was used, with “1 out of 5” being represented as .2, and “5 out of 5” represented by 1. This number served as the level of prediction for each respondent. The equation for accuracy of prediction is  $A=S/P$ , where A=accuracy of prediction, S=IQ score, and P=Level of prediction. The accuracy of prediction score will range between .2 and 5. Estimated IQ Test results will be listed as a whole number between zero and one, and determined by the number of correct answers out of the five questions.

## **Results**

### **Hypothesis #1**

The essential variables for hypothesis #1 are the accuracy of prediction value, with means ranging from -1 to 1, and the IQ score ranging from 0 to 1. A mean score on the accuracy of prediction variable of zero represents a high or perfect level of accuracy by the respondent in

predicting their actual IQ score on the measure. As shown in Table 1, the results produced a mean accuracy of prediction score of (M=.14, SD=.34) A mean score on the IQ score variable of one represents a high or perfect score on the IQ test. The results showed a mean of (M=.43, SD=.30). A Pearson bivariate correlation was run on the two variables accuracy of prediction and IQ score. As shown in Table 2, the two variables showed a strong negative correlation,  $r(22) = -.839, p < .000$ . The null hypothesis for hypothesis #1 is rejected, and support for the alternate hypothesis is found.

Table 1

*Accuracy of Prediction and IQ Score*

	N	Mean	Std. Deviation
IQ Score	22	.43	.30
Prediction Index	22	.14	.3

Table 2

*Pearson Bivariate Correlation*

	IQ Score	Prediction Index
IQ Score Pearson Correlation	1	-.839**
Sig. (2-tailed)		.000
N	22	22

\*\* Correlation is significant at the 0.01 level (2-tailed)

**Hypothesis #2**

The essential variables for hypothesis #2 are the survey version, and accuracy of prediction. The data was assessed using an independent samples t test comparing the mean accuracy of prediction in version one (no warning) and version two (warning regarding overconfidence). The mean score for the no warning condition was (N=7, M=.1470, SD=.3426). There was a slight decrease in the mean accuracy of prediction in the warning condition (N=15, M=.1329, SD=.3480) There were no significant findings regarding this interaction,  $t(22)=.089$ ,  $p = .561$ . It is worthy of noting that the randomization called for by the software did not produce evenly balanced groups in version one and version two by group size, with seven participants in the Warning condition, and 15 in the no warning condition. Level of prediction, differentiated here from the accuracy of prediction index used previously in both hypothesis, displayed the opposite results from what was predicted with significant results showing an opposite effect of the warning message on the level of prediction, with a decrease in the mean level of prediction in the warning group compared to the no warning group. The warning group having (N=15, M=.6533, SD=.1187) and a lower mean value in the No Warning condition of (N=7, M=.5714, SD=.2430). After running an independent samples T-test on level of prediction in the warning and no warning condition, the results were significant at  $t(22) = -1.07$ ,  $p = <.05$  (.018)

### **Discussion**

Support for the first hypothesis was established in finding a strong negative correlation between the IQ score of a survey respondent, and that respondent's accuracy of prediction index. These finding align with the work by Zelle & Krizan (2014) regarding self-assessments of ability in comparison to objective performance. The findings add to the support established by Dunning et al. (2004, 2018) that show a correlation between level of general ability and the accuracy of predictions regarding that ability, with lower levels of overconfidence generally associated with

higher levels of performance. Where prior research has focused on other forms of assessments to test and establish support for findings regarding overconfidence, the support found here in the use of a syllogism test of general ability is believed to be the first experiment to find a similar correlation through the use of this measure, the BAROCO short, widening the support for this hypothesis.

The use of an accuracy of prediction index, that assessed the respondent's prediction both prior to and throughout the entire assessment, helped to improve the internal and external validity of the study, meet APA standards of research, as well as control for potential confounding impacts of relying simply on one stated opportunity to capture the level of confidence, as is prevalent in a majority of the research discussed here (Association, 2019; Dunning et al. 2004, 2018; Freund & Kasten 2012; Logg et al. 2018; Zell & Krizan 2014). Where relying on an initial prediction could limit the interpretation of results, in regards to the confounds impacting prediction such as the respondents prior knowledge or familiarity with the measure, ambiguity regarding the proposed assessment or personal IQ; sampling the respondents predictions throughout the assessment, as accomplished here, attempts to produce higher external validity. In general, the methodology used to produce support for hypothesis one utilized randomization of the prediction response answer order to control for ordering effects that could have impacted the results. Additionally, the utilization of a timed assessment, and the use of the features inherent to the qualtrics software that assessed the amount of time spent on each question helped provide evidence that the respondents included in the results accurately completed the survey. These measures, combined with the design of the assessment in accordance with APA guidelines ensured maintenance of required standards of research, adding to the credibility of these findings (Association, 2019).

The second hypothesis sought to observe the impact of a warning dissuading overconfidence, on the level of overconfidence displayed by respondents. There was a slight reduction effect of the warning on the accuracy of prediction score, confirming the prediction that a warning used as a benchmark to prime respondents would decrease overconfidence. However, these results were not significant, and do not provide support for the second hypothesis. Where Dunning et al. (2004) was able to implement a benchmarking event at the onset of their survey, the benchmarking or warning message used in this survey did not produce similar results. Additionally, when comparing the level of prediction scores across the warning and no-warning condition, an opposite yet significant effect was observed where the warning condition was associated with higher levels of overconfidence, which is exactly opposite of what was predicted, possibly indicating methodological, or other research design and analysis errors.

The results of this study are limited in their external validity as the only participants were graduate students currently enrolled in psychology courses at SNHU. Likely expanding this study to participants not currently enrolled in graduate level psychology courses could improve the external validity of the study. Because of the relatively small number of participants,  $N=22$ , and the limitation to only students enrolled in a psychology course, the impacts of priming through exposure to similar topics, exposure to the actual research used to develop this study, as well as exposure to discussions posted by the author on SNHU discussion boards as part of regular course work, could impact the reliability of the study as well as the external validity. However, although these limitations could impact some of the external validity, the randomized sample achieved, methodology utilized, and analysis tools employed combined with the strength of statistical significance found in support of hypothesis number one are believed to outweigh the

limitations, and therefore the support for hypothesis number one is believed to add to the body of research that exists regarding the relationship between ability level and accuracy of predictions.

The second hypothesis, for which support was not established, highlights methodological issues that could be improved upon for future iterations of this study. The use of the Qualtrics software built in random assignment of respondents to control groups did not function as predicted. The randomization assigned only seven respondents to the no warning, or control group, and 15 respondents to the warning, or experimental group. Although the statistical analysis utilized attempts to control for unbalanced groups, in future iterations of this experiment, random assignment that results in two groups that are more equally balanced on respondent size would be desired. Additionally, in regards to the second hypothesis, the design of the cautionary warning dissuading overconfidence could likely impact the results and should be assessed in future iterations. Likely the use of multiple benchmarks or primers during the introduction to the survey would help to show whether or not the type of warning or benchmark used impacts the accuracy of prediction. As only one condition, the warning condition, which utilized the same cautionary warning statement throughout the condition, it is not possible to tell from this studies design whether or not a different form of warning or benchmarking would result in the same insignificant results.

Where Dunning et al. (2004,2018), identified that overconfidence in self assessments is a prevalent phenomenon, and Zelle & Krizan, (2012) showed that measures of general ability are included in the assessments capable of displaying significant overconfidence from respondents, this study that utilizes the BAROCO short both affirms those authors work, as well as adds a new measure with it's own unique cultural controls, and measures of objectiveness, capable of widening the understanding of this unique phenomena. With the overwhelming amount of

research conducted regarding overconfidence, (Dunning et al. 2004, 2018; Freund & Kasten 2012; Logg et al. 2018, Zell & Krizan 2014), the second hypothesis explored in this study for which further development is needed, points to the direction of future research related to this topic. Applying our understanding of overconfidence to real world problems, or to the design of future measures that wish to control for overconfidence effects in their results, is a likely next step. The second hypothesis of this study attempts to implement procedures that could be utilized to reduce the potentially confounding or negative impacts of overconfidence in self assessments, as assessed by Dunning et al. (2004, 2018). Furthermore, a deeper look into the motivations behind overconfidence is likely warranted.

Predictions regarding the motivation of individuals as displayed in survey based research, as opposed to everyday life, are likely very applicable to a conversation regarding overconfidence. Do we consistently observe overconfidence in self assessments because of an inherent desire to maintain a high self image inferred from the assumption of external validity of the survey/respondent method (Logg et al., 2018), or is it due to the method itself, and inherent limitations, in addition to the artificial nature of the survey to respondent relationship when compared to respondent to researcher relationship. Here, I predict that further analysis into the role of the survey itself, in capturing overconfidence can help to better explain these motivations, and will likely necessitate a comparison between human to human predictions and human to survey predictions. The impacts of social interaction could likely moderate and mediate the prevalence of overconfidence in ways not observable through the use of a survey administered by a computer, or even in writing without the direct interaction of a human researcher or confederate.



## References

- Adams, R. L., Smigielski, J., & Jenkins, R. L. (1984). Wechsler Adult Intelligence Scale-Revised--Short Form. *PsycTESTS*. <https://doi-org.ezproxy.snhu.edu/10.1037/t04392-000>
- Association, A. P. *Publication Manual of the American Psychological Association*®. (2019). Retrieved from <https://mbsdirect.vitalsource.com/#/books/9781433813757/>
- Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed Self-Assessment: Implications for Health, Education, and the Workplace. *Psychological Science in the Public Interest*, 5(3), 69–106. <https://doi-org.ezproxy.snhu.edu/10.1111/j.1529-1006.2004.00018.x>
- Dunning, D., Heath, C., & Suls, J. M. (2018). Reflections on self-reflection: Contemplating flawed self-judgments in the clinic, classroom, and office cubicle. *Perspectives on Psychological Science*, 13(2), 185–189. <https://doi-org.ezproxy.snhu.edu/10.1177/1745691616688975>
- Ethical principles of psychologists and code of conduct. (2002). *The American Psychologist*, 57(12), 1060–1073. Retrieved from <https://search-ebshost-com.ezproxy.snhu.edu/login.aspx?direct=true&db=cmedm&AN=12613157&site=eds-live&scope=site>
- Freund, P. A., & Kasten, N. (2012). How smart do you think you are? A meta-analysis on the validity of self-estimates of cognitive ability. *Psychological Bulletin*, 138(2), 296–321. <https://doi-org.ezproxy.snhu.edu/10.1037/a0026556>
- Heine, S. J., Lehman, D. R., Markus, H. R., & Kitayama, S. (1999). Is there a universal need for positive self-regard? *Psychological Review*, 106(4), 766–794. <https://doi-org.ezproxy.snhu.edu/10.1037/0033-295X.106.4.766>
- Logg, J. M., Haran, U., & Moore, D. A. (2018). Is Overconfidence a Motivated Bias?

Experimental. *JOURNAL OF EXPERIMENTAL PSYCHOLOGY-GENERAL*, 147(10), 1445–1465. <https://doi-org.ezproxy.snhu.edu/10.1037/xge0000500>

Shikishima, C., Yamagata, S., Hiraishi, K., Sugimoto, Y., Murayama, K., & Ando, J. (2011). A simple syllogism-solving test: Empirical findings and implications for g research. *Intelligence*, 39(2–3), 89–99.

<https://doi-org.ezproxy.snhu.edu/10.1016/j.intell.2011.01.002>

Shikishima, C., Yamagata, S., Hiraishi, K., Sugimoto, Y., Murayama, K., & Ando, J. (2011). BAROCO Short. PsycTESTS. [https://doi-org.ezproxy.snhu.edu/Full text; 999915934\\_full\\_001.pdf](https://doi-org.ezproxy.snhu.edu/Full%20text%3B%20999915934_full_001.pdf)

Zell, E., & Krizan, Z. (2014). Do people have insight into their abilities? A metasynthesis. *Perspectives on Psychological Science*, 9(2), 111–125. <https://doi-org.ezproxy.snhu.edu/10.1177/1745691613518075>