# MEASURING SCIENTIFIC WISDOM

A Thesis

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Master of Arts

by

Chak Haang Wong

Dec 2020

## ABSTRACT

The current study developed measures of scientific wisdom as practiced within the field of general sciences. A total of 125 Cornell undergraduates participated in the study, and we examined the relationship among the following scales: Scientific Reasoning, Scientific Wisdom, Scientific Creativity, a typical-performance scale of self-assessed wisdom scale (SAWS), and three psychometric tests. The results showed that the first three scales formed one clear principal component while the psychometric tests of intelligence, SAT and ACT scores, and GPA formed another, which suggested that the scientific tests we created assessed abilities that were different from general intelligence (*g*). This finding implied that colleges and universities that are seeking future scientists would want to consider supplementing the *g*-based conventional tests with tests that measure skills related to scientific reasoning, scientific wisdom, and scientific creativity.

Keywords: scientific wisdom, scientific reasoning tests, scientific creativity, college admissions, generating hypotheses, designing experiments, drawing conclusions, general intelligence, cognitive scales, general sciences.

# BIOGRAPHICAL SKETCH

Chak Haang Wong has lived in Shenzhen, Hong Kong, and Shanghai, where she completed her elementary school, middle school, and high school education, respectively. Prior to enrolling at Cornell as a sophomore, Chak Haang was a student at the College of San Mateo. She served as a student senator and international student ambassador who has planned, organized, and evaluated campus events; she was an academic advisor and student mentor for the incoming students. As an undergraduate at Cornell University, Chak Haang has been a mentor of Cornell College Mentor for Kids, a mentor of Cornell Peer Partnership, and an activity leader of Cornell Summer Program for English and Culture. She has also been a teaching assistant of Dr. Robert J. Sternberg in two courses, The Nature of Human Intelligence and Intimate Relationship: Liking, Loving, and Interpersonal Attraction. In terms of research, Chak Haang has worked on two projects measuring Scientific Reasoning skills, two projects measuring Scientific Wisdom, and two projects measuring Cultural Intelligence. She has also created the general-sciences related to Scientific Reasoning scales, the Scientific Wisdom scale, and half of the Cultural Intelligence test. Chak Haang has published one study on the topic of Scientific Reasoning (Sternberg, Wong, and Sternberg, 2019), and she has been working with Dr. Robert J. Sternberg on another possible publication. Chak Haang Wong received her Bachelor of Science degree in Human Development at Cornell University. She is currently a master's student in Developmental Psychology at Cornell University and is planning on pursuing a Ph.D. degree in the near future.

ACKNOWLEDGEMENTS

This thesis could not have been completed without the guidance from my faculty advisors, assistance from the Cornell Statistical Consulting Unit, and the participation of Cornell students. My deepest appreciation

I would like to express the deepest appreciation to my faculty advisor and committee chair, Dr. Robert J. Sternberg, who has always supported me and patiently provided editing and comments on my thesis. I am very honored to be his research assistant since sophomore year, as well as a teaching assistant since a senior year at Cornell University. Starting from a beginner who has little knowledge about the field of intelligence, Dr. Sternberg has taught me with his immense knowledge, helpful responses, and generous encouragements to guide me view intelligence more than just IQ scores and cognitive abilities – intelligence should also include creativity, adaptation, practical skills, and more importantly, wisdom. I am thankful for his guidance and opinion as now I view the world a little bit differently, and I feel empowered to make the world a better place.

I would like to express my gratitude to my secondary advisor Dr. Lin Bian for her supports and feedback on my thesis throughout the year. I would also like to thank Françoise Vermeylen from the Cornell Statistical Consulting Unit, as well as Dr. Karin Sternberg, for providing me help with analyzing data on SPSS.

Last but not least, my special thanks go to my parents and all my friends for their continuous supports and love throughout my master's program.

TABLE OF CONTENTS

LIST OF TABLES

## Measuring Scientific Wisdom

Wisdom has become increasingly crucial in our society today. Brienza, Kung, Santos, Bobocel, and Grossmann (2017) have defined wisdom as the ability "to discern how to act in a particular situation, with an aim to achieve a situation-appropriate balance between different moral virtues and personal preferences." Wise people also show several common characteristics, including (1) balance of alternative viewpoints, (2) epistemic humility, (3) contextual adaptability, and (4) multiple perspectives on problems (Grossmann et al., 2020).

Although the four viewpoints mentioned above are alleged to represent a consensus among experts, Sternberg (2020) has pointed out that the common good and the ethical element seem to be missing. The world is not in lack of smart people but wise people, who could genuinely make a difference, with the infusion of positive ethical values, by applying critical thinking and combining factual knowledge to design solutions for creating a common good over the short and long term (Sternberg, 1998, 2019).

When the world confronts difficult times like the COVID-19 crisis, there are groups of intellectuals that shall contribute their knowledge to developing vaccines and cures for the virus; however, the majority of people require wisdom to protect themselves. Wisdom can be seen through correctly distinguishing false and misleading information on social media and breaking news or through finding ways to establish connections and continue activities virtually that were once in-person. It is wisdom that can sustain people's lives and allow them to adapt to these evolving circumstances, especially in difficult times.

More importantly, with the rise of STEM-related disciplines, scientific innovations are now more widespread than ever. The scientific innovations, such as genetic editing, artificial intelligence, big-data analysis, and antibiotics, although they have been developed using the accumulation of factual knowledge and have achieved unprecedented development, can have negative implications that require more thinking and consideration (Sternberg, 2020). Signs have shown how artificial intelligence can surpass humans in certain behaviors, whether it is physical tasks of packing machines in factories or mental analysis of beating human opponents in chess games (Koch, 2015). The birth of the first genetically-modified twins resulting from gene-editing by a Chinese scientist caused widespread consternation and sparked intense moral arguments (Li, Walker, Nie, & Zhang, 2019). Big-data analysis has permeated every bit of our lives, manipulating our preferences and privacy (Kruse, Goswamy, Raval, & Marawi, 2016). Antibiotics have been proven helpful in killing bacteria, but overusing antibiotics can instead be harmful (Mayo Clinic Staff, 2020).

In other words, we do not have sufficient *scientific wisdom* in tackling our own creations – we still lack certain factors of wisdom, especially the awareness, management, and reflectivity to evaluate these inventions. Without scientific wisdom, these technologies could be put to dangerous use, including terrorist attacks and political conspiracies, and we are at risk of losing our control over our own scientific innovations.

Despite being so essential to the modern world, wisdom has not been developed enough in the education system (Sternberg, 2001). In a conventional societal setting, schools are more often regarded as solely the place to teach the "hard," or solid knowledge and facts, which in turn equips students for entering the existing power structure of the society, as well as selects its "finest" based in part on how much knowledge the students acquire. It is hard to incorporate teaching for wisdom into schools because the schools heavily emphasize general intelligence (g) and performance on conventional academic tests.

Sternberg and Sternberg (2017), as well as Sternberg, Wong, and Sternberg (2019), created Scientific Reasoning scales based on Sternberg's theory of successful intelligence. The theory suggested that the three components of intelligence – analytical, creative, and practical – function collectively within a sociocultural context to achieve adaptation to the environment (Sternberg, 1988, 1997, 1999). With this being said, students need not only analytical intelligence (which are close to the concept of general intelligence in conventional academic settings) but also creative intelligence and practical intelligence to excel in doing sciences. For example, students who scored highly on *g*-based academic tests, such as the SAT and GRE, may not be able to apply these analytical skills to their work life and personal life (Sternberg 1985, 2005).

The two studies, Sternberg and Sternberg (2017) and Sternberg, Wong, and Sternberg (2019), created the Scientific Reasoning scales. The first two subtests of the Scientific Reasoning scales, Generating Hypotheses and Designing Experiments, measured creative intelligence; the third subtest, Drawing Conclusions, measured analytical intelligence; and as a whole, the test also measured practical intelligence, or in other words, the practice of doing science. The two studies explored the relationship between the scientific reasoning scales and tests of general intelligence (g), such as the SAT, ACT, and two psychometric tests of inductive reasoning. The results showed that scores on the three subtests positively (and significantly) correlated with each other but were weakly (and in some cases, negatively) correlated with students' SAT, GPA, and psychometric test scores.

With the theory of successful intelligence in mind, the current study not only adopted the Scientific Reasoning scales but also created the Scientific Wisdom scale to incorporate the "wisdom" part of intelligence and further the previous studies' applications. Scientific wisdom is different from scientific knowledge, analytical reasoning, creativity, and practical skills. Through creating the Scientific Wisdom scale, we aimed to examine students' ability to consider reflectively and deeply about the short- and long-term positive and negative implications of scientific innovations. To assess creativity, we also used the Scientific Creativity Scale, in which students were asked to propose any research questions and design an empirical study to answer the questions. Moreover, the widely used Self-Assessed Wisdom Scale (SAWS) was included in this study to add a multidimensional perspective of wisdom. According to Webster (2003), the five dimensions of wisdom, Experience, Openness, Emotional Regulation, Reminiscence and Reflectiveness, and Humor, are interdependent and cohesive that all are considered necessary components of wisdom.

As scientists are living in a hyper-competitive environment and are often encouraged to create scientific innovations, how can we build safeguards into STEM research and development and measure, or even teach, scientific wisdom? The goal of the current study was to investigate empirically how scientific wisdom can be understood and measured.

Similar to the previous studies, the current study aims examined the relationship of the Scientific Wisdom scale to the Scientific Reasoning scales, the Scientific Creativity scale, the Self-Assessed Wisdom Scale (SAWS), as well as the psychometric tests (Letter Sets, Figure Classification, and Word Beginnings Test).

It is predicted that the Scientific Wisdom scale will correlate significantly with the Scientific Reasoning scales, the Scientific Creativity scale, and the SAWS, but not as highly with the psychometric scales, as well as GPA, SAT, and ACT scores.

## Method

### Participants

A total of 125 Cornell undergraduate students participated in the data collection. In all, 86 of them were female, and 39 of them were male. The average age of the participants was 19.51 years, with a standard deviation of 1.21. Students were enrolled in 28 different majors, with 7 students not indicating their majors and 4 students reporting double majors. The racial composition was 45.2% White or Caucasian, 26.6% Asian, 8.9% Black or African American, 8.1% Hispanic or Latino, 3.2% American Indian or Alaska Native, 0.8% Native Hawaiian or Other Pacific Islander, 6.5% two or more races, and 0.8% preferring not to answer.

**Materials**

Six assessments were arranged in the form of a questionnaire, in the following order: psychometric assessments, which included a letter-sets test, a figure-classification test, and a vocabulary test; a scientific-wisdom test; scientific-reasoning tasks, which included three sections (generating hypotheses, generating experiments, and drawing conclusions); a scientific-creativity test; a Self-Assessed Wisdom Scale (SAWS); and a demographic questionnaire. The scientific-wisdom test was informally piloted on a small sample of students before being used in the study.

The psychometric scales were adapted from Kit of Factor-Referenced Cognitive Tests (Ekstrom, French, Harman, & Dermen, 1976). They measured fluid and crystallized intelligence. They were recorded and measured based on the number of correct answers, and each correct answer was granted one point.

**Letter Sets.** The Letter-Sets test had 15 problems, and each problem contained five sets of four letters. We used Part 1 of the Letter Sets Test from the Kit of Factor-Referenced Cognitive Tests. The task was to cross out one set of letters that was different than the rest of the sets. For example, among NOPQ, DEFL, ABCD, HIJK, and UVWX, DEFL should be crossed out because it is the only set of letters that is not in alphabetical order. The time limit for this assessment was 7 minutes. In previous research, we corrected for guessing, but the corrected scores were almost perfectly correlated with the uncorrected scores, so we did not correct them here.

**Figure Classification.** The Figure-Classification tested the participants' ability to discover rules that explain things. We used Part 1 of the Figure-Classification test from the Kit of Factor-Referenced Cognitive Tests, and there was a total of 14 problems. In each problem, there were either two or three groups, each consisting of three figures. The participants were asked to look for property that is the same among the three figures in any one group, as well as for property that makes the groups differ from one another. For example, it is given that Group 1 has shaded squares and Group 2 has unshaded squares, and when a line of figures was presented, the participants need to write "1" under each figure that has a shaded square as in Group 1, and "2" under each figure with an unshaded square as in Group 2. The time limit for this assessment was 8 minutes.

**Vocabulary.** The Vocabulary test assessed the participants' knowledge of word meanings. We used Level 3 (median level) of the vocabulary test from the Kit of Factor-Referenced Cognitive Tests. There were 24 problems. In each problem, a word was given, for example, the word "jovial," and the participants were asked to choose one of the five numbered words that have the same meaning or nearly the same meaning as the word given (the correct answer would be "jolly" instead of "refreshing," "scare," "thickset," or "wise). The time limit for this assessment was 6 minutes.

**Scientific Wisdom.** The Scientific Wisdom test was created for the current study. There were 5 items that measured student's scientific wisdom as experienced within the field of general sciences. For each of the items, the participants read a short passage describing some innovative scientific products or novel scientific ideas. Then, they were asked to list as many questions and/or comments to examine/criticize the societal and scientific costs and benefits of the innovations. One point was granted for each valid comment or question, and it was scored based on (1) analytical strength, (2) usefulness and practicality, (3) creativity, and (4) wisdom toward serving a common good. An example of a nutritional supplementation passage would be:

"A scientific entrepreneur has developed a nutritional supplement that he believes offers great promise for helping people to lose weight. Because it is a supplement to be sold over the counter, it is not subject to the FDA (Food and Drug Administration) approval. He has tested the nutritional supplement on 200 people, half of whom were randomly assigned to the supplement condition and half to a placebo condition. People, of course, did not know to which condition they were assigned. After 3 months in which members of each group took their respective daily pills, the experimental (supplement) group had lost an average of 21 pounds, whereas the placebo group had lost an average of 1 pound. A venture-capital firm is considering offering the entrepreneur $5.5 million to start producing, marketing, and selling the product. You have been asked to advise the venture-capital firm as to whether to fund the production, marketing, and sales of the supplement. What would you want to ask and/or tell them?" There was no time limit for this assessment. Please see Appendix A for a complete set of items.

For each section of the scientific-reasoning tasks, we adapted three items in connection with general science and related disciplines from Sternberg, Wong, and Sternberg (2019). There was a total of 15 items. In terms of scoring, for Generating Hypotheses, one point was granted for a valid and reasonable hypothesis. For Generating Experiments and Drawing conclusions, each item was judged on a 5-point scale (from 1 = poorly answered to 5 = thoroughly answered). There was no time limit for this assessment. Please see Appendix B for a complete set of items.

**Generating Hypotheses.** A short scenario was described, and the participants were asked to generate as many alternative hypotheses as they could at the moment. For example:

"Jasper is interested in the function of water when growing plants. He adds 50 mL of water to the earth in which half of his plants are growing and 150mL of water to the earth in which the other half of his plants are growing. He notices that plants with the 150mL of water grow taller than those with the 50 mL of water and claims that water helps the plants to grow even more. What are some alternative hypotheses regarding why the plants with 150mL of water grow taller?"

**Generating Experiments.** A short description of a hypothesis was presented, and the participants were asked to design an experiment that could adequately test the hypothesis. For example:

"Ivan is interested in the relationship between environment and mental health. He believes that environments that are loud and noisy create stress for people, and thus lead to mental- health issues. However, Ivan does not know how to test this hypothesis. Please suggest an experimental design to test this hypothesis and describe the experiment in some detail. Assume you have the resources you need to be able to do the experiment."

**Drawing Conclusions.** A short description of an experiment's results was presented, and the participants were asked to explain whether or not the conclusions drawn from the results were valid. For example:

"Anna was interested in how well roses grow if she waters them with different kinds of vitamin supplements mixed in the water. She watered half of her roses with diluted liquid

Vitamin B and the other half with diluted liquid Vitamin E. The results showed that roses watered with liquid Vitamin B grew faster than roses watered with liquid Vitamin E. She concluded that Vitamin B is more beneficial than Vitamin E for plant growth. Is the conclusion correct? Why or why not?"

**Scientific Creativity.** The Scientific-Creativity scale asked the participants to propose a research question about human behavior and then to design an empirical study to answer the question. The response was judged on a 5-point scale (1 = poorly answered to 5 = thoroughly answered), and there was no time limit for this assessment. Scoring was particularly attuned to the scientific creativity of the response. Please see Appendix C for a description of the item.

**Self-Assessed Wisdom Scale (SAWS).** The SAWS was adapted from Webster (2003). We used the first 20 items. The questionnaire asked participants to indicate their level of agreement on a 6-point scale (from 1 = strongly disagree to 6 = strongly agree) to a series of statements reflecting five factors of personal experience. The five factors were experience, emotional regulation, reminiscence and reflection, humor, and openness. There were 4 statements for each factor. For example, the statement "I have overcome many painful events in my life" would correspond to the "experience" factor, whereas "It is easy for me to adjust my emotions to the situation at hand" would correspond to the "emotional regulation" factor. The scoring of this scale was the summed total score of the 20 statements. There was no time limit for this assessment. Please see Appendix D for a complete set of items.

**Demographic Questionnaire.** A demographic questionnaire assessed variables including gender, age, year at Cornell, major, GPA, SAT, and ACT scores, socially-defined race, number of lab course taken, number of math and science courses taken, whether or not the participant has taken research-methods courses and if so, how many, and whether or not the participant was contemplating a career in a science-related field.

### Design

In the analysis of variance, the dependent variables were standardized test scores, and the independent variables were participant sex and race. In the correlational analyses, the dependent variables were standardized test scores, and the independent variables were the Scientific-Reasoning Test scores, the Scientific-Wisdom Test score, the Scientific-Creativity score, and the SAWS score.

### Procedure

The participants were first asked to read and sign the informed consent form. Then, the researcher handed out a set of assessments in the form of surveys to the participants. The assessments were arranged in the following order: Letter Sets, Figure Classification, Vocabulary, Scientific Wisdom, Generating Hypotheses, Generating Experiments, Drawing Conclusions, Scientific Creativity, the Self-Assessed Wisdom Scale, and a brief demographic questionnaire. The first three assessments had a seven-minute, eight-minute, and six-minute time limit, respectively, and they were timed by an experimenter with a stopwatch. After completion of the Vocabulary test, the participants were asked to read the instructions carefully and then to finish the rest of the surveys at their own pace. After completion of the demographic questionnaire, the researcher collected the survey and presented a debriefing form of the study. The entire session lasted about 1.5 hours.

## Results

The report of results is divided into four parts: basic statistics, analysis of variance, correlations, and factor analyses.

**Basic Statistics**

Table 1 shows the means and standard deviations of all the measures used in the study. Cronbach alpha reliabilities of the scales were 0.67 for Letter Sets, 0.95 for Figure Classification, 0.71 for Vocabulary, 0.90 for Scientific Wisdom, 0.79 for Generating Hypothesis, 0.60 for Generating Experiments, 0.43 for Drawing Conclusions, and 0.74 for the combined scientific-reasoning scales (Generating Hypotheses, Generating Experiments, Drawing Conclusions, combined). The Scientific Creativity test did not yield internal-consistency reliability because it comprised only one item.

Because the number of participants who took either/both SAT and ACT was not sufficient for reliable data analysis, we converted all the self-reported SAT total scores (SAT reading score plus SAT math score) to ACT scores into the "SAT/ACT" variable. We used the "ACT to SAT Score Conversion Chart" by the Princeton Review. For those who reported both SAT and ACT scores, an average was calculated.

**Table 1.** Mean Scores and Standard Deviations.

| | **Mean** | **Standard Deviation** | **N** |
|---|---|---|---|
| Age | 19.51 | 1.21 | 125 |
| Cumulative College GPA | 3.53 | 0.35 | 84 |
| SAT Reading Score | 703.90 | 55.33 | 77 |
| SAT Math Score | 752.73 | 59.18 | 77 |
| ACT Score | 32 | 2.98 | 66 |
| SAT/ACT | 31.7 | 3.42 | 118 |
| Letter Sets total score | 11.17 | 2.38 | 125 |
| Figure Classification total score | 73.56 | 17.24 | 125 |
| Vocabulary total score | 10.26 | 3.67 | 125 |
| Scientific Wisdom total score | 22.78 | 8.15 | 124 |
| Hypotheses total score | 7.70 | 2.77 | 125 |
| Experiments total score | 8.40 | 1.66 | 124 |
| Conclusions total score | 6.18 | 1.77 | 125 |
| Scientific Creativity | 3.42 | 0.81 | 119 |
| SAWS total score | 88.28 | 9.74 | 125 |

Note: SAWS refers to Selthe f-Assessed Wisdom Scale.

**Analysis of Variance**

Table 2 shows the ANOVA table for gender differences by comparing the means, and Table 3 shows the ANOVA table for ethnicity differences by comparing the means. Because there were no significant sex or ethnicity differences on none of the test measures, no follow-ups are reported.

**Table 2.** Gender Analysis of Variance.

| | | | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|---|
| GPA * Gender | Between Groups | (Combined) | 0.000 | 1 | 0.000 | 0.000 | 0.993 |
| | Within Groups | | 9.972 | 82 | 0.122 | | |
| | Total | | 9.972 | 83 | | | |
| SAT Reading score * Gender | Between Groups | (Combined) | 295.196 | 1 | 295.196 | 0.095 | 0.758 |
| | Within Groups | | 232335.973 | 75 | 3097.813 | | |
| | Total | | 232631.169 | 76 | | | |
| SAT Math score * Gender | Between Groups | (Combined) | 139.942 | 1 | 139.942 | 0.039 | 0.843 |
| | Within Groups | | 265987.330 | 75 | 3546.498 | | |
| | Total | | 266127.273 | 76 | | | |
| ACT score * Gender | Between Groups | (Combined) | 0.067 | 1 | 0.067 | 0.007 | 0.932 |
| | Within Groups | | 575.933 | 64 | 8.999 | | |
| | Total | | 576.000 | 65 | | | |
| Letter Sets total score * Gender | Between Groups | (Combined) | 2.726 | 1 | 2.726 | 0.480 | 0.490 |
| | Within Groups | | 698.746 | 123 | 5.681 | | |
| | Total | | 701.472 | 124 | | | |
| Figure Classification total score * Gender | Between Groups | (Combined) | 13.228 | 1 | 13.228 | 0.044 | 0.834 |
| | Within Groups | | 36843.572 | 123 | 299.541 | | |
| | Total | | 36856.800 | 124 | | | |
| Vocabulary total score * Gender | Between Groups | (Combined) | 92.072 | 1 | 92.072 | 7.158 | 0.008 |
| | Within Groups | | 1582.216 | 123 | 12.864 | | |
| | Total | | 1674.288 | 124 | | | |
| Scientific Wisdom total score * Gender | Between Groups | (Combined) | 44.573 | 1 | 44.573 | 0.670 | 0.415 |
| | Within Groups | | 8120.548 | 122 | 66.562 | | |
| | Total | | 8165.121 | 123 | | | |
| Hypotheses total score * Gender | Between Groups | (Combined) | 4.967 | 1 | 4.967 | 0.644 | 0.424 |
| | Within Groups | | 949.081 | 123 | 7.716 | | |
| | Total | | 954.048 | 124 | | | |
| Experiments total score * Gender | Between Groups | (Combined) | 3.538 | 1 | 3.538 | 1.291 | 0.258 |
| | Within Groups | | 334.300 | 122 | 2.740 | | |
| | Total | | 337.839 | 123 | | | |
| Conclusions total score * Gender | Between Groups | (Combined) | 8.583 | 1 | 8.583 | 2.777 | 0.098 |
| | Within Groups | | 380.185 | 123 | 3.091 | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| | Total | | 388.768 | 124 | | | |
| Scientific Creativity * Gender | Between Groups | (Combined) | 0.001 | 1 | 0.001 | 0.001 | 0.975 |
| | Within Groups | | 76.991 | 117 | 0.658 | | |
| | Total | | 76.992 | 118 | | | |
| SAWS total score * Gender | Between Groups | (Combined) | 1.378 | 1 | 1.378 | 0.014 | 0.905 |
| | Within Groups | | 11759.822 | 123 | 95.608 | | |
| | Total | | 11761.200 | 124 | | | |

**Table 3.** Ethnic-Group Analysis of Variance.

| | | | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|---|
| GPA * Ethnicity | Between Groups | (Combined) | 1.059 | 7 | 0.151 | 1.290 | 0.267 |
| | Within Groups | | 8.913 | 76 | 0.117 | | |
| | Total | | 9.972 | 83 | | | |
| SAT Reading score* Ethnicity | Between Groups | (Combined) | 21945.134 | 7 | 3135.019 | 1.027 | 0.421 |
| | Within Groups | | 210686.035 | 69 | 3053.421 | | |
| | Total | | 232631.169 | 76 | | | |
| SAT Math score* Ethnicity | Between Groups | (Combined) | 70758.674 | 7 | 10108.382 | 3.570 | 0.002 |
| | Within Groups | | 195368.599 | 69 | 2831.429 | | |
| | Total | | 266127.273 | 76 | | | |
| ACT * Ethnicity | Between Groups | (Combined) | 82.251 | 6 | 13.708 | 1.638 | 0.153 |
| | Within Groups | | 493.749 | 59 | 8.369 | | |
| | Total | | 576.000 | 65 | | | |
| Letter Sets total score * Ethnicity | Between Groups | (Combined) | 52.351 | 7 | 7.479 | 1.348 | 0.234 |
| | Within Groups | | 649.121 | 117 | 5.548 | | |
| | Total | | 701.472 | 124 | | | |
| Figure Classification total score * Ethnicity | Between Groups | (Combined) | 5543.634 | 7 | 791.948 | 2.959 | 0.007 |
| | Within Groups | | 31313.166 | 117 | 267.634 | | |
| | Total | | 36856.800 | 124 | | | |
| Vocabulary total score * Ethnicity | Between Groups | (Combined) | 72.017 | 7 | 10.288 | 0.751 | 0.629 |
| | Within Groups | | 1602.271 | 117 | 13.695 | | |
| | Total | | 1674.288 | 124 | | | |
| Scientific Wisdom total score * Ethnicity | Between Groups | (Combined) | 339.164 | 7 | 48.452 | 0.718 | 0.657 |
| | Within Groups | | 7825.957 | 116 | 67.465 | | |
| | Total | | 8165.121 | 123 | | | |
| Hypotheses total score* Ethnicity | Between Groups | (Combined) | 56.193 | 7 | 8.028 | 1.046 | 0.403 |
| | Within Groups | | 897.855 | 117 | 7.674 | | |
| | Total | | 954.048 | 124 | | | |
| Experiments total score * Ethnicity | Between Groups | (Combined) | 37.609 | 7 | 5.373 | 2.076 | 0.052 |
| | Within Groups | | 300.230 | 116 | 2.588 | | |
| | Total | | 337.839 | 123 | | | |
| Conclusions total score * Ethnicity | Between Groups | (Combined) | 11.550 | 7 | 1.650 | 0.512 | 0.824 |
| | Within Groups | | 377.218 | 117 | 3.224 | | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | Total | | 388.768 | 124 | | | |
| Scientific Creativity * Ethnicity | Between Groups | (Combined) | 1.264 | 7 | 0.181 | 0.265 | 0.966 |
| | Within Groups | | 75.728 | 111 | 0.682 | | |
| | Total | | 76.992 | 118 | | | |
| SAWS total score * Ethnicity | Between Groups | (Combined) | 760.170 | 7 | 108.596 | 1.155 | 0.334 |
| | Within Groups | | 11001.030 | 117 | 94.026 | | |
| | Total | | 11761.200 | 124 | | | |

**Correlations**

Table 4 shows a complete correlation table for all the variables in this study. The results supported our prediction in the ways that (a) the three subtests of the Scientific Reasoning scales (Generating Hypotheses, Generating Experiments, and Drawing Conclusions) positively and significantly correlated with each other; (b) the Scientific Reasoning scales also positively and significantly correlated with the Scientific Wisdom scale and the Scientific Creativity scale; (c) the Self-Assessed Wisdom Scale (SAWS) positively and significantly correlated with the Scientific Wisdom scale and the Scientific Creativity scale; (d) none of the above-mentioned scales significantly correlated with the psychometric scales (Letter Sets, Figure Classification, Vocabulary); in some cases, the scientific-thinking scales negatively correlated with the psychometric scales; (e) none of the above-mentioned scales (except for Generating Hypotheses) significantly correlated with either SAT scores or ACT score; (f) all the psychometric scales positively and significantly correlated with students' GPA, SAT Reading and Math scores, and ACT.

**Table 4.** Correlations

| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Gender | 1 | 0.043 | 0.069 | -0.001 | -0.036 | 0.023 | -0.011 | -0.071 | 0.128 | -0.020 | -0.062 | -0.019 | .235** | 0.074 | 0.072 | -0.102 | -0.149 | -0.003 | 0.011 |
| 2 | Age | 0.043 | 1 | .875** | -0.053 | -0.006 | -0.104 | 0.115 | -0.022 | .352** | .320** | 0.048 | -0.011 | 0.026 | 0.151 | .240** | -0.122 | 0.035 | 0.017 | 0.084 |
| 3 | Year | 0.069 | .875** | 1 | 0.080 | 0.022 | -0.078 | 0.153 | -0.044 | .502** | .409** | 0.108 | -0.019 | 0.128 | 0.133 | .227* | -0.145 | 0.078 | -0.004 | 0.006 |
| 4 | GPA | -0.001 | -0.053 | 0.080 | 1 | .417** | .322* | .316* | .228* | 0.188 | -0.025 | .276* | 0.132 | 0.010 | 0.214 | 0.200 | 0.154 | 0.136 | -0.085 | -0.184 |
| 5 | SAT Reading | -0.036 | -0.006 | 0.022 | .417** | 1 | .462** | .810** | .691** | 0.044 | 0.065 | 0.186 | 0.021 | .333** | 0.057 | .258* | 0.123 | 0.003 | -0.014 | 0.106 |
| 6 | SAT Math | 0.023 | -0.104 | -0.078 | .322* | .462** | 1 | .597** | .680** | 0.174 | 0.060 | .274* | 0.164 | 0.081 | 0.127 | 0.201 | 0.210 | 0.130 | -0.141 | 0.060 |
| 7 | ACT | -0.011 | 0.115 | 0.153 | .316* | .810** | .597** | 1 | .721** | 0.149 | 0.035 | 0.172 | 0.089 | .333** | 0.023 | .301* | 0.003 | -0.046 | -0.121 | 0.077 |
| 8 | SAT/ACT | -0.071 | -0.022 | -0.044 | .228* | .691** | .680** | .721** | 1 | 0.021 | 0.063 | .232* | 0.140 | 0.145 | -0.026 | .217* | 0.040 | -0.025 | -0.090 | 0.065 |
| 9 | STEM Courses | 0.128 | .352** | .502** | 0.188 | 0.044 | 0.174 | 0.149 | 0.021 | 1 | .541** | .192* | 0.120 | 0.105 | 0.149 | 0.133 | -0.139 | -0.003 | -0.167 | -0.074 |
| 10 | Lab Courses | -0.020 | .320** | .409** | -0.025 | 0.065 | 0.060 | 0.035 | 0.063 | .541** | 1 | .205* | 0.156 | 0.055 | 0.165 | 0.002 | -.184* | 0.076 | -0.122 | -0.008 |
| 11 | LS | -0.062 | 0.048 | 0.108 | .276* | 0.186 | .274* | 0.172 | .232* | .192* | .205* | 1 | .298** | 0.138 | 0.104 | 0.126 | -0.032 | 0.052 | -.212* | -0.023 |
| 12 | FC | -0.019 | -0.011 | -0.019 | 0.132 | 0.021 | 0.164 | 0.089 | 0.140 | 0.120 | 0.156 | .298** | 1 | 0.115 | -0.051 | -0.007 | -.264** | -0.115 | -0.164 | -0.084 |
| 13 | Vocabulary | .235** | 0.026 | 0.128 | 0.010 | .333** | 0.081 | .333** | 0.145 | 0.105 | 0.055 | 0.138 | 0.115 | 1 | 0.024 | 0.126 | 0.013 | -0.115 | 0.023 | 0.009 |
| 14 | Scientific Wisdom | 0.074 | 0.151 | 0.133 | 0.214 | 0.057 | 0.127 | 0.023 | -0.026 | 0.149 | 0.165 | 0.104 | -0.051 | 0.024 | 1 | .461** | .389** | .324** | .328** | .199* |
| 15 | Hypotheses | 0.072 | .240** | .227* | 0.200 | .258* | 0.201 | .301* | .217* | 0.133 | 0.002 | 0.126 | -0.007 | 0.126 | .461** | 1 | .323** | .292** | .233* | -0.006 |
| 16 | Experiments | -0.102 | -0.122 | -0.145 | 0.154 | 0.123 | 0.210 | 0.003 | 0.040 | -0.139 | -.184* | -0.032 | -.264** | 0.013 | .389** | .323** | 1 | .457** | .410** | -0.019 |
| 17 | Conclusions | -0.149 | 0.035 | 0.078 | 0.136 | 0.003 | 0.130 | -0.046 | -0.025 | -0.003 | 0.076 | 0.052 | -0.115 | -0.115 | .324** | .292** | .457** | 1 | .415** | 0.104 |
| 18 | Scientific Creativity | -0.003 | 0.017 | -0.004 | -0.085 | -0.014 | -0.141 | -0.121 | -0.090 | -0.167 | -0.122 | -.212* | -0.164 | 0.023 | .328** | .233* | .410** | .415** | 1 | .207* |
| 19 | SAWS | 0.011 | 0.084 | 0.006 | -0.184 | 0.106 | 0.060 | 0.077 | 0.065 | -0.074 | -0.008 | -0.023 | -0.084 | 0.009 | .199* | -0.006 | -0.019 | 0.104 | .207* | 1 |

*. Correlation is significant at the 0.05 level (2-tailed).
**. Correlation is significant at the 0.01 level (2-tailed).
Note: LS refers to Letter Sets, FC refers to Figures Classification, SAWS refers to Self-Assessed Wisdom Scale.

## Factor Analyses

Both varimax-rotated principal-components and principal-factor matrices were also used for the analyses. The factor analyses helped to further evaluate the interrelations between the scales.

Tables 5 and 6 show the varimax-rotated principal-component and principal-factor matrices for the psychometrics scales, the Scientific Wisdom scale, the Scientific Reasoning scales, the Scientific Creativity scale, and the SAWS. Similar to the correlational analyses, the results from both tables supported our predictions and showed that our Scientific Reasoning scales, Scientific Wisdom scale, and Scientific Creativity scale loaded on a different factor from either the cognitive scales (Letter Sets, Figure Classification, and Vocabulary) or the SAWS. All the cognitive scales loaded on one factor, and the SAWS loaded on its own factor, perhaps because it was the only typical-performance self-report measure.

**Table 5.** Rotated Principal-Components Matrix: Cognitive scales, scientific wisdom scale, scientific reasoning scale, scientific creativity scale, and SAWS[a].

| | Component | | |
|---|---|---|---|
| | 1 | 2 | 3 |
| Letter Sets | 0.11 | 0.71 | -0.24 |
| Figure Classification | -0.20 | 0.69 | -0.05 |
| Vocabulary | 0.01 | 0.51 | 0.28 |
| Scientific Wisdom | 0.74 | 0.13 | 0.20 |
| Hypotheses | 0.71 | 0.29 | -0.03 |
| Experiments | 0.76 | -0.25 | -0.13 |
| Conclusions | 0.72 | -0.14 | 0.02 |
| Scientific Creativity | 0.57 | -0.28 | 0.40 |
| SAWS | 0.04 | -0.02 | 0.89 |

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
[a] Rotation converged in 4 iterations.

Notes: Three principal components had Eigenvalues greater than 1. Component 1 had an Eigenvalue of 2.62, accounting for 29.1% of the variance in the data. Component 2 had an Eigenvalue of 1.48, accounting for 16.4% of the variance in the data. Component 3 had an Eigenvalue of 1.07, accounting for 11.8% of variance in the data. Cumulative percent variance accounted for was 57.3%.

**Table 6.** Rotated Principal-Factor Matrix: Cognitive scales, scientific wisdom scale, scientific reasoning scale, scientific creativity scale, and SAWS[a].

| | Factor | | |
|---|---|---|---|
| | 1 | 2 | 3 |
| Letter Sets | 0.05 | 0.53 | -0.11 |
| Figure Classification | -0.19 | 0.49 | -0.03 |
| Vocabulary | 0.02 | 0.22 | 0.01 |
| Scientific Wisdom | 0.67 | 0.14 | 0.23 |
| Hypotheses | 0.61 | 0.26 | 0.01 |
| Experiments | 0.73 | -0.26 | -0.20 |
| Conclusions | 0.60 | -0.10 | 0.09 |
| Scientific Creativity | 0.51 | -0.28 | 0.26 |
| SAWS | 0.08 | -0.07 | 0.57 |

Extraction Method: Principal Axis Factoring.

Rotation Method: Varimax with Kaiser Normalization.

[a] Rotation converged in 4 iterations.

Notes: Three principal factors had Eigenvalues greater than 1. Component 1 had an Eigenvalue of 2.62, accounting for 29.1% of the variance in the data. Component 2 had an Eigenvalue of 1.48, accounting for 16.4% of the variance in the data. Component 3 had an Eigenvalue of 1.07, accounting for 11.8% of variance in the data. Cumulative percent variance accounted for was 57.3%.

Tables 7 and 8 show the varimax-rotated principal-component and principal-factor matrices for the Scientific Wisdom scale, Scientific Reasoning scales, Scientific Creativity scale, SAWS, and the *g*-based college-admission tests. Similar to Tables 5 and 6, the results show that the students' SAT/ACT scores loaded on one factor, and SAWS loaded on a second factor. The Scientific Wisdom scale, Scientific Reasoning scales, and Scientific Creativity scale loaded on a third factor, suggesting they measure highly related skills.

**Table 7.** Rotated Principal-Component Matrix: Scientific wisdom scale, scientific reasoning scales, scientific creativity scale, SAWS, and college-Admissions Tests (*g*-based)[a].

| | Component | | |
|---|---|---|---|
| | 1 | 2 | 3 |
| Scientific Wisdom | 0.71 | 0.18 | 0.11 |
| Hypotheses | 0.64 | -0.09 | 0.51 |
| Experiments | 0.78 | -0.11 | 0.01 |
| Conclusions | 0.75 | 0.08 | -0.09 |
| Scientific Creativity | 0.62 | 0.43 | -0.24 |
| SAWS | 0.03 | 0.95 | 0.09 |

| | | | |
|---|---|---|---|
| SAT/ACT | -0.07 | 0.07 | 0.90 |

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
[a] Rotation converged in 6 iterations.

Notes: Three principal components had Eigenvalues greater than 1. Component 1 had an Eigenvalue of 2.55, accounting for 36.4% of the variance in the data. Component 2 had an Eigenvalue of 1.15, accounting for 16.4% of the variance in the data. Component 3 had an Eigenvalue of 1.07, accounting for 15.3% of variance in the data. Cumulative percent variance accounted for was 68.1%.

**Table 8.** Rotated Principal-Factor Matrix: Scientific wisdom scale, scientific reasoning scales, scientific creativity scale, SAWS, and college-Admissions Tests (*g*-based)[a.]

| | Factor | | |
|---|---|---|---|
| | 1 | 2 | 3 |
| Scientific Wisdom | 0.55 | 0.28 | 0.14 |
| Hypotheses | 0.46 | 0.86 | -0.07 |
| Experiments | 0.70 | 0.03 | -0.10 |
| Conclusions | 0.68 | 0.00 | 0.04 |
| Scientific Creativity | 0.62 | -0.08 | 0.26 |
| SAWS | 0.09 | 0.06 | 0.79 |
| SAT/ACT | -0.06 | 0.25 | 0.04 |

Extraction Method: Principal Axis Factoring.
Rotation Method: Varimax with Kaiser Normalization.
[a] Rotation converged in 5 iterations.

Notes: Three principal factors had Eigenvalues greater than 1. Component 1 had an Eigenvalue of 2.55, accounting for 36.4% of the variance in the data. Component 2 had an Eigenvalue of 1.15, accounting for 16.4% of the variance in the data. Component 3 had an Eigenvalue of 1.07, accounting for 15.3% of variance in the data. Cumulative percent variance accounted for was 68.1%.

## Discussion

The current study sought to address the questions, “What is scientific wisdom?”, “Why is it important?” and “How can it be measured?” We developed a measure of scientific wisdom as practiced within the field of general sciences, and we examined its relationship with the Scientific Reasoning Scales, the Scientific Creativity Scale, the Self-Assessed Wisdom Scale (SAWS), and the *g*-based psychometric tests (Letter Sets, Figure Classification, and Vocabulary). We also asked the participants to report their cumulative GPA, SAT Reading and Math scores, and ACT scores for correlational analyses.

The results of this study aligned with our hypotheses: The Scientific Wisdom scale correlated significantly with the Scientific Reasoning scales, the Scientific Creativity scale, and the SAWS, but not with the psychometric tests or conventional academic tests. This pattern of results suggested that our tests of scientific skills seem to be measuring abilities that are different from what the conventional *g*-loaded academic tests are measuring.

As knowledge in science and technology are vastly growing in modern times, more and more students are majoring in the sciences and working hard to become successful scientists. College and universities thus should be cultivating and fostering students' abilities to do "hands-on" scientific research (and the accompanying scientific-reasoning skills), to generate creative innovations (scientific creativity), and to use scientific knowledge to serve a common good (scientific wisdom). To develop scientists competent to do actual research in science, not merely succeed in taking courses in science, universities need to go beyond emphasizing high GPA or scores on conventional *g*-based academic tests.

The findings of this study will also contribute to the scientific study of wisdom, which is a comparably less well-developed field than that of the study of general intelligence. Our Scientific Wisdom scale aims to measure the students' ability to think about the short/long-term positive and negative implications of scientific innovations for the common good, which is different from scientific reasoning and creativity. Good scientists should not only be able to propose creative research questions and perform valid experiments; they also, perhaps even more importantly, should be able to examine/criticize the societal and scientific costs and benefits of scientific innovation, based on the consideration of positive ethical and moral values.

The results of the study suggest that a valid test of scientific wisdom may be a useful supplement to conventional academic tests, such as SAT, ACT, or GRE, in school admissions, especially in the fields of sciences. Moreover, for students who are contemplating a career in general sciences and related disciplines, such a test could encourage students to develop scientific wisdom.

Like any study, the current study had some limitations. First, we did not have a large degree of diversity in terms of participants' educational background because our participants were all undergraduate students at a highly selective university. Second, we do not know whether or not the use of our test of scientific wisdom predicts students' performance in university training or in an actual scientific career. Third, the Self-Assessed Wisdom Scale (SAWS) failed to correlate positively with the Scientific Reasoning scales and loaded on its own factor in the factor analyses, indicating a possible limitation of the use of such a self-report scale, and instead, a more performance-based scale could be used in further studies.

To conclude, the current study showed a significant relationship between the Scientific Wisdom scale, the Scientific Reasoning scales, and the Scientific Creativity scale. More importantly, the relationship between these scales and the psychometric scales, as well as GPA, SAT, and ACT scores, was relatively weak. We hope that emphasis on scientific wisdom and on implementing a Scientific Wisdom scale might lead future generations to become better scientists and thinkers who make the right decisions.

We are living in a time of extreme scientific challenges—a pandemic, air and water pollution, global climate change, and increasing resistance of bacteria to antibiotics. If ever there has been a time to teach science students to think not only logically but also wisely, this would seem to be it. We believe that developing scientific wisdom and assessing it would represent a step in the right direction.

# APPENDIX

## Appendix A. Scientific Wisdom Scale

*For each of the following items, you will be reading about some innovative products and novel scientific ideas. As with any innovation, there are potential pluses and minuses to each idea. Thus, despite their benefits, there might be drawbacks.* *<u>Please note that there is no time limit.</u>*

*You will be asked to examine/criticize the costs and benefits of the studies by generating as many questions and/or comments as you can. (Please keep in mind that the studies described are hypothetical. None has actually been done or is being planned to be done.)*

*Questions/comments will be scored based on (1) The number of relevant questions/comments provided, (2) Analytical strength, (3) Usefulness and practicality, (4) Creativity, and (5) Wisdom toward serving a common good.*

*<u>Example:</u>*

**Nutritional supplementation.** A scientific entrepreneur has developed a nutritional supplement that he believes offers great promise for helping people to lose weight. Because it is a supplement to be sold over the counter, it is not subject to FDA (Food and Drug Administration) approval. He has tested the nutritional supplement on 200 people, half of whom were randomly assigned to the supplement condition and half to a placebo condition. People of course did not know to which condition they were assigned. After 3 months in which members of each group took their respective daily pills, the experimental (supplement) group had lost an average of 21 pounds whereas the placebo group had lost an average of 1 pound. A venture-capital firm is considering offering the entrepreneur $5.5 million to start producing, marketing, and selling the product. You have been asked to advise the venture-capital firm as to whether to fund the production, marketing, and sales of the supplement. What would you want to ask and/or tell them?

<u>*Possible Questions/Comments*</u>:

1. Have potential side effects of the weight-reducing supplement been adequately studied, or studied at all?

2. Are there any undesirably consequences of using the drug, such as physical or psychological addiction?

3. For what population or populations is the supplement safe? Children, and if so, of what ages? Older people? People with weakened immune systems?

4. Does the supplement interact with pharmaceutical drugs, such that there might be unexpected side effects for people taking those drugs, including weight-loss drugs?"

5. Maybe 3 months was not long enough to assess the supplement's effectiveness. The supplement's effect might decrease over time.

6. Can the supplement safely be used by people with unusual weight-related conditions, such as morbid obesity or anorexia (i.e., people taking the supplement to lose weight that they do not need to lose)?

7. Can the supplement be priced so that it is available to all who really would benefit from it, or will it "price out" many people who potentially could use it but who could not afford it?"

8. Should this substance be sold over the counter?

1. **Fossil Plant.** Data indicated that residents who are living close to a coal-fired power station showed a higher rate of asthma, heart disease, and low birth weight; air pollution apparently caused by the power station is a risk factor for all these conditions. Because of this, the coal company plans to retire the power station. After assessing multiple factors, such as

employment levels and air/water quality, researchers found that retirement of the power station would create the greatest reduction in negative environmental effects on health; however, it would also greatly negatively impact the employment and local economy. Moreover, losing jobs could affect health through increased risk of heart attack, alcohol or drug abuse, and decreased mental health status and physical activity. You have been asked to advise the fossil company as to how to proceed. What would you want to ask and/or tell them?

2. **Air Pollution.** The NYC Department of Public Health received an energy-efficiency program proposal designed to reduce air pollution exposure. The program researchers have shown that, in many urban areas, significant disparities in urban air pollution exposures often are associated with significant health consequences. For example, low-income and vulnerable populations often live in places with unhealthful exposures to environmental factors that lead to higher risks of heart diseases. The proposed program aimed to minimize the health impact of urban air pollution by mainly targeting indoor air quality. The program proposed to perform energy upgrades (for example, adding an indoor filtered mechanical ventilation system) that would improve energy efficiency in residential homes. The upgrade would cost up to $11,000 per home, to be paid for by the homeowners; however, this price could be prohibitive for low-income families. You have been asked to advise the Department of Public Health as to whether the proposal is potentially feasible and effective, and whether they should endorse the program. What would you want to ask and/or tell them?

3. **Copper Mining.** Scientists have found that copper is not only a highly utilized metal for electrical and household objects but also an effective antimicrobial surface. Bacteria, such as Salmonellae and Cronobacter sakazakii, often found in food contaminations, can be rapidly killed with copper alloys. Scientists then suggested that copper-containing solutions could be

used in the healthcare environment (such as hospitals and clinics) and could be applied to prevent/kill bacterial and fungal infections. A copper company has been contacted by scientists and is interested in purchasing over 10,000 square miles of land for copper mining and making more copper-containing solutions. However, copper mining negatively impacts the environment and creates sources of pollution. You have been asked to advise the company as to whether to purchase the lands. What would you want to ask and/or tell them?

4. **Genetic Alteration.** A genetic engineer has recently invented a type of chemical fertilizer to grow more “advanced” agricultural products, such as sweeter tomatoes, bigger watermelons, and herbicide-resistant crops. Successful completion of his invention has produced food and crops that are bigger and grow faster, that are more resistant to disease, and that thrive in different environments. More importantly, this invention could reduce levels of hunger throughout the world. Even though the scientist assured potential investors that the breeding process of the products was only an extension of the “natural” way of breeding, there are potential side effects caused by the chemicals. For examples, chemical fertilizers may lead to fewer or poorer quality crops in the long term because of the intricacies of soil health, excessive use of chemicals in the soil can kill off fish in nearby bodies of water, and so on. The scientist asked for a company to fund his invention. You have been asked to advise the company as to whether to fund the invention. What do you want to ask and/or tell them?

5. **Antibiotic resistance.** Scientists in a pharmaceutical firm have developed a new antibiotic that treats a range of respiratory illnesses of bacterial origin. In clinical tests, the drug has shown itself to be effective and to be superior to alternative medications in that bacteria have not developed resistance to it. After a long process, the drug has won FDA approval, which is hard to get. At the same time, the firm realizes that doctors and patients alike are often reluctant

to be among the first to try out a new antibiotic drug. The company has received a large order from a country that it knows routinely allows antibiotics to be sold without a doctor's prescription. The company recognizes that selling antibiotics in this way has disadvantages, such as patients excessively buying/taking the drug without knowing whether they are allergic to it or whether the particular antibiotic is right for the particular infection. However, this way of selling would allow the company to recover the tremendous cost of developing the medication, as well as the need to show domestic markets that the drug is being used effectively. You have been asked to advise the pharmaceutical firm as to whether to fulfill the order. What would you want to ask and/or tell them?

## Appendix B. Scientific Reasoning Scales

### Generating Hypotheses

1. Mary is interested in the effects of breakfast on academic performance and behavior of adolescents. One day, she eats a small piece of steak and two eggs for breakfast, and she feels focused when she does a mock SAT test early in the morning. She starts to think that breakfast is beneficial and adolescents who eat a breakfast high in protein daily will perform behaviorally and academically better in school.

*What are some alternative hypotheses regarding why Mary is focused when she does the mock SAT test?*

2. Jasper is interested in the function of water when growing plants. He adds 50 mL of water to the earth in which half of his plants are growing and 150mL of water to the earth in which the other half of his plants are growing. He notices that plants with the 150mL of water grow taller than those with the 50 mL of water and claims that water helps the plants to grow even more.

*What are some alternative hypotheses regarding why the plants with 150mL of water grow taller?*

3. James is interested in the effect of drinking tea and coke. He drinks 150 mL of green tea every day for two weeks and then after that, he drinks 150 mL of coke for another two weeks. He finds no significant change in his health during the times when he drinks green tea only or coke only. He claims that coke is not as harmful as people say it would be.

*What are some alternative hypotheses regarding why there is no change in James' health when he does the experiment?*

**Generating Experiments**

1. Ivan is interested in the relationship between environment and mental health. He believes that environments that are loud and noisy create stress for people, and thus lead to mental- health issues. However, Ivan does not know how to test this hypothesis.

*Please suggest an experimental design to test this hypothesis and describe the experiment in some detail. Assume you have the resources you need to be able to do the experiment.*

2. Jon believes that radiation from electronic devices such as cell phones and computers negatively influences people's health, especially when the people are at a younger age. However, Jon is not sure how to test this hypothesis.

*Please suggest an experimental design to test this hypothesis and describe the experiment in some detail. Assume you have the resources you need to be able to do the experiment.*

3. April is interested in making her seeds sprout faster. She hypothesizes that higher temperatures can make the seeds sprout faster and earlier. However, April is not sure how to properly design an experiment to test this hypothesis.

*Please suggest an experimental design to test this hypothesis and describe the experiment in some detail. Assume you have the resources you need to be able to do the experiment.*

**Drawing Conclusions**

1. Henry tested the hypothesis that sugar helps students to be alert and focus in class. Three groups of 200-participants each were asked to consume a piece of dark chocolate (~8 g sugar), or milk chocolate (~10 g sugar), or completely sugar-free chocolate. It was found that both groups that consumed sugar were more alert and focused in class than was the sugar-free group; moreover, the milk chocolate group performed better than the dark chocolate group in paying attention. He concluded that students should eat snacks that contain more sugar in order to stay awake in class.

*Is the conclusion correct? Why or why not?*

2. Tony wanted to see whether exercising for 30 minutes per day would reduce cholesterol levels. He had 10 subjects exercise for at least 30 minutes per day for three weeks and 10 subjects not exercise at all. He also controlled the number of calories all subjects consumed daily. He found that cholesterol levels were reduced for those who did the exercise. He concluded that exercising reduces cholesterol levels and thus lowers the risk of having heart disease.

*Is the conclusion correct? Why or why not?*

3. Anna was interested in how well roses grow if she waters them with different kinds of vitamin supplements mixed in the water. She watered half of her roses with diluted liquid Vitamin B and the other half with diluted liquid Vitamin E. The results showed that roses watered with liquid Vitamin B grew faster than roses watered with liquid Vitamin E. She concluded that Vitamin B is more beneficial than Vitamin E for plant growth.

*Is the conclusion correct? Why or why not?*

## Appendix C. Scientific Creativity Scale

What is a study about human behavior that you might like to design and conduct? What is a question about human behavior that you consider important that you would like to answer? How might you answer it through research? Please write your response below. <u>Please note that there is no time limit.</u>

## Appendix D. Self-Assessed Wisdom Scale (SAWS)

This brief questionnaire is designed to investigate how people of different ages perceive themselves with respect to life experiences and whether or not these perceptions change as we grow older. You are asked to rate all of the following statements using the scale below. Remember, there are no “right” or “wrong” answers and your responses will remain anonymous. Do not rush but work steadily as we are interested in your first impressions. Please record your responses by writing only one number on the rating scale to the right of each statement.

1 = Strongly Disagree; 2 = Moderately Disagree; 3 = Slightly Disagree

4 = Slightly Agree; 5 = Moderately Agree; 6 = Strongly Agree

1. I have overcome many painful events in my life.
2. It is easy for me to adjust my emotions to the situation at hand.
3. I often think about connections between my past and present.
4. I can chuckle at personal embarrassments.
5. I like to read books which challenge me to think differently about issues.
6. I have had to make many important life decisions.
7. Emotions do not overwhelm me when I make personal decisions.
8. I often think about my personal past.

9. There can be amusing elements even in very difficult life situations.

10. I enjoy listening to a variety of musical styles besides my favourite kind.

11. I have dealt with a great many different kinds of people during my lifetime.

12. I am “tuned” in to my own emotions.

13. I reminisce quite frequently.

14. I try and find a humorous side when coping with a major life transition.

15. I enjoy sampling a wide variety of different ethnic foods.

16. I am very good at reading my emotional states.

17. Reviewing my past helps me gain perspective on current concerns.

18. I am easily aroused to laughter.

19. I often look for new things to try.

REFERENCES

Brienza, J. P., Kung, F. Y. H., Santos, H. C., Bobocel, D. R. R., & Grossmann, I. (2017). Wisdom, bias, and balance: Toward a process-sensitive measurement of wisdom-related cognition. *Journal of Personality and Social Psychology*, *115,* 1093-1126.

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). Manual for kit of factor-referenced cognitive tests. *Princeton, NJ: Educational testing service*.

Grossmann, I., Weststrate, N. M., Ardelt, M., Brienza, J. P., Dong, M., Ferrari, M., Fournier, M. A., Hu, C. S., Nusbaum, H. C., Vervaeke, J. (2020). The science of wisdom in a polarized world: Knowns and unknowns. *Psychological Inquiry,31,* 103-133.

Li, J. R., Walker, S., Nie, J. B., & Zhang, X. Q. (2019). Experiments that led to the first gene-edited babies: the ethical failings and the urgent need for better governance. *Journal of Zhejiang University. Science. B*, *20*, 32 – 38.

Koch, C. (2015). Will artificial intelligence surpass our own? *Scientific American Mind 26,* 26-29.

Kruse, C. S., Goswamy, R., Raval, Y., & Marawi, S. (2016). Challenges and opportunities of big data in health care: A systematic review. *JMIR medical informatics*, *4*, 38.

Sternberg, R. J. (1985). Human abilities: An information-processing approach. *W.H. Freeman: New York.*

Sternberg, R. J. (1988). The triarchic mind: A new theory of human intelligence*. New York: Viking*.

Sternberg, R. J. (1997). Thinking styles*. New York: Cambridge University Press*.

Sternberg, R.J. (1998). A balance theory of wisdom*. Rev. Gen. Psychol. 2,* 347-365.

Sternberg, R.J. (1999). Successful intelligence: Finding a balance. *Trends Cognit*, *3*, 436–442.

Sternberg, R. J. (2001). Why schools should teach for wisdom: The balance theory of wisdom in educational settings. *Educational Psychologist, 36,* 227-245.

Sternberg, R.J. (2005). The domain generality versus specificity debate: How should it be posed? *Creativity Across Domains: Faces of the Muse,* 299-306.

Sternberg, R. J. (2019). Why people often prefer wise guys to guys who are wise: An augmented balance theory of the production and reception of wisdom. In *Cambridge handbook of wisdom;* Sternberg, R. J., Glueck J. Eds.; Cambridge University Press, 162-181.

Sternberg, R.J.; Wong, C.H.; Sternberg, K (2019). The relation of tests of scientific reasoning to each other and to tests of general intelligence. *Journal of Intelligence, 7,* 20.

Sternberg, R. J. (2020). The missing links: Comments on "the science of wisdom in a polarized world." *Psychological Inquiry, 31*, 153-159.

Mayo Clinic Staff (2020). Antibiotics: Are you misusing them? *Mayo Clinic.* https://www.mayoclinic.org/healthy-lifestyle/consumer-health/in-depth/antibiotics/art-20045720.

Webster, J. D. (2003). An exploratory analysis of a self-assessed wisdom scale. *Journal of Adult Development, 10,* 13-22.