



Minnesota State University, Mankato
Cornerstone: A Collection of Scholarly
and Creative Works for Minnesota
State University, Mankato

All Graduate Theses, Dissertations, and Other
Capstone Projects

Graduate Theses, Dissertations, and Other
Capstone Projects

2017

Evaluating a Measure of Student Effectiveness in an Undergraduate Psychology Program

Colin Omori
Minnesota State University, Mankato

Follow this and additional works at: <https://cornerstone.lib.mnsu.edu/etds>

 Part of the [Higher Education Commons](#), and the [Other Psychology Commons](#)

Recommended Citation

Omori, C. (2017). Evaluating a Measure of Student Effectiveness in an Undergraduate Psychology Program [Master's thesis, Minnesota State University, Mankato]. Cornerstone: A Collection of Scholarly and Creative Works for Minnesota State University, Mankato. <https://cornerstone.lib.mnsu.edu/etds/719/>

This Thesis is brought to you for free and open access by the Graduate Theses, Dissertations, and Other Capstone Projects at Cornerstone: A Collection of Scholarly and Creative Works for Minnesota State University, Mankato. It has been accepted for inclusion in All Graduate Theses, Dissertations, and Other Capstone Projects by an authorized administrator of Cornerstone: A Collection of Scholarly and Creative Works for Minnesota State University, Mankato.

Evaluating a Measure of Student Effectiveness in an Undergraduate Psychology Program

by Colin Omori

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Arts in
Industrial-Organizational Psychology at Minnesota State University - Mankato

Date 5/9/2017

This thesis paper has been examined and approved.

Examining Committee

Dr. Andrea Lassiter Chairperson

Dr. Emily Stark

Dr. Paul Mackie

Abstract

This study evaluates the utility of a measure of student knowledge in various areas of psychology. The 40-item measure is a revision of a pilot test from last year, and was distributed to 90 undergraduate psychology students at a Midwestern university. The average score on the assessment was 19.87 (SD = 6.20), or 49.68%. Future directions for the assessment include validation by professors, leading to the revision and removal of items. An additional suggestion is increased enforcement in regard to student participation: The study's 90-student sample was obtained after omitting 173 students who did not finish the exam, did not consent, or finished in an amount of time deemed impossible by the researchers.

Introduction

Institutions of higher learning across the country value self-assessment for a multitude of reasons; accrediting bodies require it, government funding depends on it (National Conference of State Legislatures, 2015), and exceptional results give them a competitive edge and an improved public image, which in turn can increase applicant pools, research output, received grants, quality of hires, and more. These forces, when considered along those outlined by Kraut & Korman (1999)--from demographic changes due to birth rates and immigration to changes brought upon the modern workforce due to new technology or new legislation--create a far more complex, dynamic, and competitive market, both in industry jobs and academia, than those in previous generations had, and thus create an impetus wherein an institution of higher learning that wishes to keep its curricula and evaluations relevant must update them fairly regularly. Updating, however, necessitates the existence of a good starting point. As such, truly assessing the effectiveness of collegiate programs of study can often prove to be difficult, particularly when taking into account the exceptionally competitive nature of today's job market that awaits graduates and the dynamic nature of our ever-evolving fields of study, both in their content and their method of delivery; simply put, effectiveness does not entail what it used to. Professors must consider what psychology means today when teaching, as opposed to what it meant in the past; it is a more socially aware, interconnected science than it used to be. The combined efforts of said professors create that which is assessed by the department, and the continuous updating of the professors' class content necessitates the continuous updating of its evaluation.

A recommendation in the APA's "Principles for Quality Undergraduate Education in Psychology" (Halpern et al, 2010) outlines the importance of the psychology department working together as a whole in order to identify and assess desired learning outcomes. The

discussion of what is important to and what is expected of the department must be discussed by those within it. The APA states, “we cannot improve undergraduate education without assessing what and how much students know when they complete our programs” (Halpern et al, 2010, p. 14).

In that case, it is necessary to tie the goals of the institution and the undergraduate major to the content of the classes themselves. In the context of the current study, I decided to focus on assessing undergraduate psychology students using a custom-developed student learning outcomes assessment test, as a piece of a comprehensive program assessment. To help guide this process, the APA outlines a lengthy list of recommendations known as the ‘Guidelines for the Undergraduate Psychology Major’ (2016). A 2013 revision of the original guidelines set forth in 2006, the ‘Guidelines 2.0’ outline a cohesive effort to “describe and develop high-quality undergraduate programs in psychology” (p. 1). They outline all-encompassing goals that an undergraduate program should provide for its students.

The first, and thus presumably the most important of these goals is a knowledge base; a solid foundation of basic psychological knowledge (key concepts, principles, etc) that individuals can draw from in problem-solving and critical analysis. This foundation must also extend to application; students must not only be able to list these concepts and principles, but apply them to real-life situations. The article goes on to mention that psychology (and really, any other field of science) is constantly evolving, particularly in the information age, and a certain flexibility—a consistent ability to revise and build upon previous knowledge—is certainly desirable. A key challenge with this is that departments and programs will need to examine, on an on-going basis, what exactly it is that students know.

This means that content knowledge of psychology will consistently changing, and assessing student effectiveness within a program of study (indeed, assessing anything) begins with content validity. That is, we must truly measure that which we aim to (in this case, a direct assessment of student knowledge gained from course content). It is described by Haynes, Richard, & Kubany (1995) as the degree to which elements of an assessment instrument are relevant to and representative of the targeted construct for a particular assessment purpose. This is distinct from face validity in that face validity is the extent to which something literally *appears* to measure a given construct purely on its surface—do the questions in a given assessment seem as if they are related to that which the assessment purports to measure (Lynn, 1986)? While face validity may seem important in assessment (Ekuma, 2012), content validity provides numerical evidence that the questionnaire which you are administering is truly representative of what you intend to measure (Lynn, 1986). In a psychology program, we would need to know that the knowledge students gain is actually the knowledge we intend for them to gain.

Furthermore, it is important to briefly discuss discriminant validity, or the extent to which a construct is distinct from other constructs (Campbell, 1960); if this measure doesn't discriminate between the knowledge that lower-division students have when compared to upper-division students, the measure cannot be called valid. Discriminant validity is what differentiates constructs, such as how different groups of people answer a questionnaire. In terms of the current study, discriminant validity will hopefully be shown between upper- and lower-division students (such that upper-division students will score higher, because they have had more time to develop a base of content knowledge in the field).

To further establish initial steps towards content and discriminant validity, a distinction must be made between performance and effectiveness. Campbell, McCloy, Oppler, and Sager (1993)

define performance as “something that people actually do and can be observed” (p. 35). It includes actions and behaviors that are relevant to the organization's goals and that can be measured in terms of each person's proficiency. They go on to say “[performance] is not the consequence or result of action, it is the action itself...” whereas effectiveness is “the evaluation of the results of performance. By definition...a measure of effectiveness is controlled by more than the actions of the individual” (p. 35). As it relates to the current study, this means student effectiveness can be thought of as the result of their actions, such as their performance on a test. Thinking a bit abstractly, a students’ ‘job,’ according to this definition, is to learn. Their performance is what they do in order to accomplish this. Their effectiveness, then, is the sum of their efforts, which in the current study will be assessed by a test. Their effectiveness (and what we ultimately aim to measure) is the result of instructors’ performance—the extent to which students remember what they are taught. With high content validity, such a measure will give us exactly that.

As noted in an article by the APA, assessing a common core of criteria within an undergraduate psychology program obviously necessitates a common core, and “there is a surprising dearth of information on this point” (2014, p. 13). Due to the wide variety of textbooks used by professors and instructors across different universities and different sections of classes, there is little cohesion across these professors, universities, and classes. Content analyses of introductory psychology textbooks have not found consistency across content, length, or comprehensiveness (Griggs, Bujak-Johnson, & Proctor, 2004; Landrum, Gurung, & Spann, 2012). Furthermore, there is little demonstrated consistency across the responses of instructors in terms of the importance of given topics (Miller & Gentile, 1998). These factors alone denote a bevy of roadblocks in accurately assessing program effectiveness on anything more than an

institution-by-institution level. In the context of the current study, our concern is to accurately measure psychology content knowledge as it is taught at this university.

The APA (2014) found only one article covering whether there was consistency in the content of introductory courses. Homa et al. (2013) saw that the majority of 158 psychology 101 syllabi made the aforementioned Principles for Quality Undergraduate Education's pillar of a knowledge base a priority, as well as emphasizing research methods and real-life application. Instructors in certain fields of psychology, however, were prone to spending disproportionately more time on their expertise than other areas of psychology, and the history of psychology, research methods, and developmental psychology tended to be underscored in favor of topics in physiological and cognitive psychology, undermining any chance of a greater degree of consistency between classes, universities, or even instructors.

Going beyond solely the content of introductory courses, Stoloff et al (2009) found that the most common classes in an undergraduate program were as follows (circa 2005), with 80% or more of studied universities containing a class on the following topics: research methods, developmental psychology, introductory psychology, abnormal psychology, social psychology, personality psychology, some form of independent study, physiological psychology, history and systems, a form of practicum or field experience, and finally, learning and memory. If some sort of assessment or even method of assessment can be considered valid across universities, it may be a better idea to begin by assessing these areas. The assessment tool designed and evaluated in this current study includes an examination of content across some common topic areas in psychology.

A survey by the Association of American Colleges and Universities found that 78% of member institutions have a "common set of intended learning outcomes for all their

undergraduate students” (p.1) but that a lack of communication is reflected in that only 42% of member institutions’ administrators believe that the majority of students understand these goals or intended outcomes.

Furthermore, only 72% of AAC&U institutions assessed their learning outcomes, although 24% say they are planning an assessment. More institutions assess at a departmental level than a general level (68% compared to 52%), although 48% assess at both levels. Only 6% of institutions did not assess or plan to assess at the time of the poll. The most common methods of assessment across these institutions are “rubrics of student work (40%), capstone projects (37%), and student surveys (35%)” (2009, p. 2).

A 2016 survey by the same association found that for those institutions assessing outcomes at the general level, rubrics applied to samples of student work made up 91% of the total. Capstone projects were identified as being 78% of the total. Additionally, at this point in time, 87% of member institutions said they were assessing outcomes across the curriculum, compared to 72% of the institutions polled in the 2009 survey.

When it comes to assessing cumulative learning outcomes, 67% of the sample said they were in this process, compared to 52% in the 2009 survey. Standardized assessments of general knowledge were used by 33% of member institutions, with an additional 38% assessing general skills, such as critical thinking (2016). The message is clear; assessing student learning outcomes is on the rise, and although standardized assessments of general knowledge are not used by the majority of polled institutions, they can be an important piece of that puzzle. In the case of the current study, this particular university is interested in this outcome, and needs a test to assess that; without knowing what students retain from their classes, an institution of higher learning cannot improve the delivery of that content.

One objective for assessing student learning outcomes outlined by the university in this study is the knowledge and application prioritized by the classes in Stoloff, et. al (2009). In this goal, students should be able to “demonstrate a broad base of knowledge of psychological science with the ability to apply within a wide variety of contexts.” Furthermore, students should apply knowledge in a manner that is ethically appropriate, socially responsible, and sensitive to populations from diverse backgrounds. The sub-goals for this include “describing key concepts, principles, and overarching themes in psychology, developing a working knowledge of psychology’s content domains, describing applications of psychology, applying ethical standards to evaluate psychological science and practice, applying psychological content and skills to career goals, and refining project management skills so as to be able to successfully complete projects in a timely manner.” These goals are primarily achieved in the classes of Introduction to Psychological Science, Careers in Psychology, and Research Methods and Design, which are courses common in most undergraduate psychology courses across the United States.

Another learning goal in this university is critical thinking; an ability to “apply scientific thinking to evaluating evidence and multiple sources of information about behavior, including being an effective consumer of information, thinking critically and analyzing the behavior and emotions of self and others, and being able to identify and interpret journal articles” attained by “engaging in innovative and integrative thinking and problem-solving, exhibiting self-efficacy and self-regulation, and developing meaningful professional direction for life after graduation.” The classes which cover this domain include Statistics for Psychology and Research Methods and Design. It is for these reasons, these goals of the university, that I developed and tested a measure of student content knowledge in psychology at the undergraduate level.

The Present Study

Questions for the measure were taken from practice questions for the psychology GRE subject test; as such, this measure is strictly meant for internal use only. The idea behind my measure is to localize in such a way that it reflects student effectiveness through the judgement of the instructors from a particular university; using the entire psychology GRE would assess areas of psychology which may not be wholly emphasized at this school. Additionally, we do not aim to assess readiness for graduate school, as in the case of the GRE. The goal here is to assess content knowledge of psychology as it is taught at this university.

Due to the proprietary nature of the measures like the GRE, psychometric data are not readily available, nor would it be applicable due to the different purposes of the two assessments. In the context of the current study, psychometric data was precisely the reason for the measure's creation. It needed to discriminate between those with no experience in psychology and those who are about to graduate with a psychology degree; it needed to accurately reflect those things which the university's instructors aim to teach in their classes; it needed to show internal consistency across groups, and it needed to demonstrate these properties through the relevant statistical data.

The main purpose of this research is to help provide additional data for a department during their program review process. In order to accomplish this a few questions must be addressed regarding the use of alternative methods. First, why not use grades as an assessment tool? There are a number of reasons: The aforementioned variation between professors' grading methods and content being a primary factor. A student may be especially proficient in certain areas of psychology, while being less proficient in others, and this could be reflected in aggregated data sets for the assessment.

Next, professors have different methods for grading. For example, some give attendance points; a student who receives a C in a class with the help of participation points cannot be said to have a greater understanding of the course content than a student who receives an A- despite never receiving any attendance points, although both students can be said to have demonstrated competency in the class.

A second question that this research can help address deals with issues of competency versus mastery: In this assessment, are we looking for a minimum score which indicates a student is competent in these areas of psychology, or do we wish to set a higher benchmark and say that they've mastered the content? With grades alone, this question cannot truly be answered, due in no small part to the variation between instructors with regard to grading methods. It can be expected that the present study may be able to better to help guide and inform academic program leaders.

Finally, why not use a practice psychology GRE? The answer here lies in three domains; content inclusion, the purpose of the test, and testing fatigue. With content inclusion, we want to make sure that any assessment that is given to a student, it is evaluative of the content they have learned; that is, it cannot include things that are not being taught at the university in question. The purpose of a psychology GRE and an assessment meant to measure student knowledge of course content at a university are two separate constructs. As mentioned previously, the idea in the current study isn't to assess readiness for graduate school; it is meant to see how much students remember about the courses they have taken as an undergraduate in psychology.

In addition, a customized assessment tool can combat testing fatigue; If an assessment is too long (an admittedly grey area, and one that can only be more thoroughly addressed with subsequent revisions of the current assessment), cognitive fatigue has been shown to increase

(Ackerman & Kanfer, 2009). The assessment has to be concise; simultaneously of adequate length and adequate comprehension. A practice psychology GRE may contain a few hundred questions. For many students not preparing to apply to graduate school, this can provoke test anxiety, as well as simply be exhausting to take.

As mentioned previously, a main purpose of this research is to help a department with its student learning outcomes assessment toward program review. Considering this, a report of what students know in various areas of psychology is a key goal. Additionally, the following hypotheses will be tested in this research:

1. A student's current year in the university will be a significant predictor of his or her test score.
2. The number of psychology classes taken by a student will be a significant predictor of his or her test score.
3. GPA will be a significant predictor of a student's test score.

Methods

Measure

The measure consisted of 40 questions taken from various content areas in psychology analogous to those shown in Stoloff et al (2009): abnormal psychology, clinical psychology, research methods, psychological statistics, history and systems of psychology, cognitive psychology, social psychology, and developmental psychology. Items were culled from various psychology content tests and randomly ordered into this assessment (see Appendix for items).

The measure was delivered to participants through the online Qualtrics survey system after participants answered affirmatively to a question concerning consent.

Participants

The assessment measure was completed by upper- and lower-division students within an undergraduate psychology program at a mid-sized Midwestern university (N = 263). Other demographic data, such as age and ethnicity, were not collected from participants. For the purposes of the study, upper-division students were operationalized as being those who have taken five or more psychology classes, whereas lower-division students were those who have taken four or fewer.

Pilot Study

The first step of this study was a pilot test, conducted in the fall and spring of 2015. 60 items were compiled into a test and distributed through Qualtrics to 58 undergraduate students. In the test, 21 items were identified as having a correct-answer percentage of less than 50%. There was a mean score on the test of 57.69%. The 40-item test in the current study was shortened from this pilot test due to the potential for excessive test length in causing cognitive fatigue (Ackerman & Kanfer, 2009). Additionally, questions from other content fields were included in light of Stoloff et al (2009). A table comparing the content of the two tests is below.

Table 1

Pilot Test and Current Test Subscales and Number of Items

Pilot Test Subscales	Number of Items	Current Test Subscales	Number of Items

Social Psychology	11	Social Psychology	10
Abnormal	10	Abnormal and Clinical	8
Biological	10	History and Systems	3
Developmental	10	Developmental	3
Cognitive	8	Cognitive	3
Research Methods and Statistics	11	Research Methods and Statistics	13

Test Administration

The revised test was distributed in Spring 2017 to both lower-division and upper-division undergraduate students at a Midwestern university through Qualtrics, primarily through two courses; History and Systems of Psychology, a capstone course for baccalaureate students; and Introduction to Psychological Science, the university's introductory psychology course. Disparities between these two groups should provide some evidence of discriminant and content validity. The full text of the test, with correct answers in bold and descriptive statistics for responses, is attached in the Appendix.

Results

Unfortunately, upon review of the data, many participants had to be cut from final analyses for a number of reasons: failure to provide consent for the study, failure to complete the study by not answering all of the items, and most worryingly, not taking the study seriously and simply providing random answers. Using Qualtrics, those who either did not consent, did not finish, or

completed the test in under twenty minutes ($N = 173$) were subject to scrutiny in this regard and eliminated, leaving a total sample of 90 students; a percentage of 34.2% of the original sample.

The average score on the assessment was a 19.87, or a 49.68% ($SD = 6.20$), and a minimum score of 6 and a maximum score of 35 out of 40 indicates that not only was there a low average score, but the distribution of scores was quite spread out. Additionally, there were a fair portion of questions that received a percentage of correct answers under 50%: Items 4, 8, 13, 14-16, 20, 22-25, 30, 34, 27, 39, and 40. However, with a low average score, this is potentially to be expected. There was a relatively even distribution of students in terms of their year at the university, as shown in Table 2.

Table 2

Percentage of Students by Declared Year

Declared Year	Percentage	Number of Students
Freshman	26.67%	24
Sophomore	14.44%	13
Junior	27.78%	25
Senior	31.11%	28

Most students reported a GPA of 3.2 or above, with 90% reporting a GPA of 2.8 or above; not surprising considering the university's minimum GPA requirement of 2.7.

Table 3

Distribution of GPA

GPA	Percentage	Number of Students
3.6-4.0	24.44%	22
3.2-3.59	37.78%	34
2.8-3.19	28.89%	26
2.4-2.79	7.78%	7
2.0-2.39	1.11%	1

The majority of students in the study had either taken one to two courses in psychology, or more than nine, showing again the purposeful split between lower-division and upper-division students.

Table 4

Number of Psychology Classes Taken by Participants

Number of Psychology Classes Taken	Percentage	Total
0	8.89%	8
1-2	33.33%	30
3-4	6.67%	6
5-6	14.44%	13
7-8	6.67%	6
9+	30%	27

Scales and Percentage of Correct Answers

Individual subscales were created for each area of psychology assessed in the test: Abnormal psychology, clinical psychology, research methods, psychological statistics, history and systems of psychology, cognitive psychology, social psychology, and developmental psychology. The subscale for abnormal and clinical psychology consisted of items 1, 2, 4-7, 9, and 17. The subscale for research methods consisted of items 3, 13, 16, 19, 21, 22, and 36. The subscale for statistics consisted of items 11-16, 19-20, and 38. The subscale for history and systems of psychology consisted of items 8, 10, and 39. The subscale for cognitive psychology consisted of items 18, 22, and 23. The subscale for social psychology consisted of items 25-32, 37, and 40. The subscale for developmental psychology consisted of items 33-35.

Table 5

Mean Percentage Correct by Subscale

Subscale	Mean percent correct
Abnormal and Clinical Psychology	57.53%
Research Methods	39.37%
Statistics	44.78%
History and Systems of Psychology	35.18%
Cognitive Psychology	41.85%
Social Psychology	49.82%
Developmental Psychology	47.41%

Additionally, items were split into those that assessed factual knowledge versus applied knowledge. Factual knowledge was identified as providing an answer to a definition or choosing from a simple list (items 1-4, 7, 10-11, 18, 24-25, 29, 31, 34-36, and 38), such as the following:

24 Neurotransmitters are released from what neuronal structure?

- Dendrite
- Myelin sheath
- Mitochondrion
- Soma
- Axon terminal**

Applied knowledge was identified as a question wherein the participant is asked to take their existing knowledge and use it to identify an answer, such as one listing symptoms and asking participants to identify the disorder the symptoms describe. (items 5, 6, 8, 9, 12-17, 19-23, 26-28, 30, 32-33, 37, 39, and 40), An example would be a ‘story problem’ such as the following:

6 Which of the following individuals is most likely to benefit from taking a medication that is a selective serotonin reuptake inhibitor (SSRI)?

- Jack, who has lost his sense of identity and has wandered from his home to a distant city
- Tyler, who believes that he has several distinct personalities
- Andrea, who hears imaginary voices telling her that she will suffer a fatal accident
- Shannon, who feels helpless, worthless, and apathetic and who thinks that her life is without meaning**
- David, who falsely believes that other people are trying to kill him.

Factual questions had an average percentage of correct answers of 54.57%, and applied questions had an average percentage of correct answers of 46.48%.

Concerning the first hypothesis, a student’s year at the university was posited to be a significant predictor of their test score. A simple linear regression was conducted to determine the strength of this relationship, and the hypothesis was shown to be correct, demonstrating differential performance between lower- and upper-division students. ($\beta = .347$, $t(90) = 3.47$, $p =$

.001). It also explained a significant proportion of variance in test scores ($R^2 = .12$, $F(1, 88) = 12.04$, $p = .001$).

Concerning the second hypothesis, the number of psychology classes taken by a student was posited as being a significant predictor of their test score. A simple linear regression was conducted to determine the strength of this relationship, and the hypothesis was shown to be correct ($\beta = .288$, $t(90) = 2.826$, $p < .01$). It also explained a significant proportion of variance in test scores ($R^2 = .083$, $F(1, 88) = 7.987$, $p < .01$).

Concerning the third hypothesis, a student's GPA was posited to be a significant predictor of their test score. A simple linear regression was conducted to determine the strength of this relationship, and the hypothesis was shown to be correct ($\beta = .213$, $t(90) = 2.04$, $p < .05$). It also explained a significant proportion of variance in test scores ($R^2 = .045$, $F(1, 88) = 4.165$, $p < .05$).

For factual questions, concerning the first hypothesis, a student's year at the university was posited to be a significant predictor of their test score regarding factual questions. A simple linear regression was conducted to determine the strength of this relationship, and the hypothesis was shown to be correct, demonstrating differential performance between lower- and upper-division students. ($\beta = .807$, $t(90) = 3.123$, $p = .002$). It also explained a significant proportion of variance in test scores ($R^2 = .10$, $F(1, 88) = 9.751$, $p = .002$).

Concerning the second hypothesis, the number of psychology classes taken by a student was posited as being a significant predictor of their test score with regard to factual questions. A simple linear regression was conducted to determine the strength of this relationship, and the hypothesis was shown to be correct ($\beta = .243$, $t(90) = 2.351$, $p < .05$). It also explained a significant proportion of variance in test scores ($R^2 = .059$, $F(1, 88) = 5.526$, $p < .05$).

Concerning the third hypothesis, a student's GPA was posited to be a significant predictor of their test score with regard to factual questions. A simple linear regression was conducted to determine the strength of this relationship, and the hypothesis was shown to be correct ($\beta = .228$, $t(90) = 2.192$, $p < .05$). It also explained a significant proportion of variance in test scores ($R^2 = .052$, $F(1, 88) = 4.804$, $p < .05$).

With regard to applied questions, concerning the first hypothesis, a student's year at the university was posited to be a significant predictor of their test score. A simple linear regression was conducted to determine the strength of this relationship, and the hypothesis was shown to be correct, demonstrating differential performance between lower- and upper-division students. ($\beta = .304$, $t(90) = 2.991$, $p < .01$). It also explained a significant proportion of variance in test scores ($R^2 = .092$, $F(1, 88) = 8.949$, $p < .01$).

Concerning the second hypothesis, the number of psychology classes taken by a student was posited as being a significant predictor of their test score with regard to applied questions. A simple linear regression was conducted to determine the strength of this relationship, and the hypothesis was shown to be correct ($\beta = .295$, $t(90) = 2.900$, $p < .01$). It also explained a significant proportion of variance in test scores ($R^2 = .087$, $F(1, 88) = 8.411$, $p < .01$).

Concerning the third hypothesis, a student's GPA was posited to be a significant predictor of their test score with regard to applied questions. A simple linear regression was conducted to determine the strength of this relationship, and the hypothesis was not supported.

Discussion

In distributing this test to undergraduate students, the goal was to assess their knowledge in various areas of psychology and across the major as a whole. The low percentage of correct

answers per subscale, the poor average score, and the unfortunate omission of nearly two-thirds of the sample are certainly limitations in analyzing if the test was successful in doing so.

In future revisions of this test, the items mentioned in the study's results that received a percentage of correct answers below 50% are recommended for removal or rephrasing, perhaps to clarify certain terms or to use possible substitute terms that professors may use in their classes.

The act of asking professors' opinions brings to mind a misstep in the current study, and a crucial future step for the assessment's development; validation. Distributing this test to professors and asking them to rate the representativeness of each question to their course content on a Likert scale is imperative. The opinions of these subject matter experts are invaluable to further establishing content validity. In order to properly validate this assessment, the test should be distributed to these experts, with a Likert scale underneath each test item; the professors can then rate how representative of their course content the items are. Those that rate below a standard can be eliminated from the final test in order to establish content validity and representation (Sireci & Geisinger, 1995). More reliable subscales can be developed as a result of this, as well.

Alternatively, the validation process could start with a large pool of potential items, grouped into their respective areas of psychology and sent to professors who teach those areas. The representative questions can then be chosen out of their pools, and added to a final assessment. The validation process could also incorporate a predictive design, surveying alumni on the directions they have gone since graduating from the university, their feelings on the knowledge assessment they had taken, and the extent to which they felt the university had met their learning goals.

Finally, a free-response section on the assessment, with questions suggested and agreed upon by professors, would demonstrate the application of student knowledge on a level that is often impossible with multiple-choice questions; a level that incorporates application and the potential to show knowledge that one has accumulated, taking parts from various domains of psychology and turning them into cohesive answers to applied questions.

Concerning the simple linear regressions that constitute the hypotheses, the news is better: Almost all hypothesized constructs proved to be statistically significant predictors of the criterion of test performance, at the overall, applied, and factual question levels. The only exception to this was that GPA was not predictive of performance on applied questions at a statistically significant level.

At the overall level, the three hypothesized predictors explained a total of nearly 25% of the variance in test scores. In this case, all hypotheses were supported. This must be taken with caution as it must be continually evaluated with subsequent revisions of the assessment; a different sample will produce a different pool of answers, and the variations between their answers will not be explained in the same proportion.

At the factual level, the three hypothesized predictors explained a total of 21.1% of the variance in test scores. In this case, all hypotheses were also supported. This must also be taken with caution as it must be continually evaluated with subsequent revisions of the assessment.

At the applied level, two of the hypothesized predictors explain a total of 17.9% of the variance in test scores. As such, two of the three hypotheses were supported.

One limitation of the study that must be addressed is the disparity between the initial sample size and the final sample size. Out of 263 undergraduate volunteers, only 90 took longer than

twenty minutes to complete the test. This would mean that on average, even the participants left over took 30 seconds to read and answer each question, an admittedly liberal estimate. Speaking solely of the participants cut from the final data set, the vast majority actually put in the effort to “complete” the test by filling it out in a matter of minutes. This shows a large degree of apathy on the part of the sample, who were likely filling out the test as quickly as they could in order to obtain the extra credit offered as compensation. This begs the question; how, in the future, can we get students to take assessments like this seriously? A potential answer may be tying their participation in the assessment to a small portion of their grade in a psychology class, or mandatory participation in taking the assessment. Of course, there will still likely be those who simply fill it out at random, but, as in the current study, those participants can be eliminated from the final data set, and would more than likely make up a smaller fraction of the sample than the current study’s 65%.

Concerning the distribution of scores on a future assessment, the ideal is something close to a negatively-skewed distribution, showing a very high percentage of students passing the test according to some standard. The question must be asked, however, of competency and mastery, and how this may tie in to potential benchmarks. This question will thus be answered in two parts.

First, there is the issue of competency versus mastery itself. Does the university want to show that its students demonstrate exceptional proficiency in the subject matter, or should the students simply meet some minimum standard? I recommend that the university take a grade of 70% to be considered as showing adequate competency in their knowledge of psychology, as a student who hypothetically receives a 70% in all of his or her classes would go on to obtain a baccalaureate degree, according to the department policy that only grades above 70% (C-) will count toward

credits in the major of psychology. As additional rationale, according to Texas Tech's assessment literature, "benchmarks are frequently set at a level that correlates to average performance, which is acceptable performance to graduate" (2010, p. 16). A program may also use existing benchmark data from their own program or a similar degree program, especially from a particularly exemplary program within the university (Hatry, van Houten, Plantz, & Greenway, 1996).

It is additionally important to keep in mind that mastery may be too high of a bar, especially in these early stages of assessment development. As an assessment is further developed, validated, and taken by more people, then perhaps it can be adjusted so as to provide a test that can be reasonably expected to be mastered (i.e., a graduating senior should be able to get a 90% or more on this), but this kind of assessment simply requires further development.

If the current goal is competency, what should be considered the proper ratio of graduating students receiving a 70% or above on this assessment? Even controlling for all other classes except the tested seniors, the mean on the current assessment is 21.88, shy of the 28 necessary for a 70% grade (provided, of course, that the content within the assessment would presumably be validated by professors). Only four tested seniors scored a 28 or higher. This question is left up to the university, but I suggest a ratio of an absolute minimum of 50% of students scoring 70% or higher on the assessment; a ratio which they will almost certainly surpass with validation, further development, and enforcement of test-taking.

With validation, we can be more confident that the content of the test is representative of the content that professors teach in their classes, and predictive of subsequent student outcomes. With further development (i.e. revision or rephrasing of inadequate questions), the subscales would likely increase in terms of their reliability, and better scores would be more likely, as well.

Finally, with more enforcement of test-taking, a larger sample size who additionally takes the test more seriously is better-assured.

The assessment of psychology content knowledge is only one piece of the puzzle when it comes to program review; nevertheless, it is a crucial piece for this university in particular, which is interested in finding out how much of students' course content they are retaining, in order to better facilitate the content and to find out what it is that they want to deliver that they can improve upon. A comprehensive program assessment should poll students and professors on the extent to which they feel they have met the goals of learning and application set by the university. The APA charts a path for psychology departments to follow (2016), and the psychology knowledge test is but one part of this process. Granted, the definition of success will and should vary from institution to institution, as each person within the department provides their insight, allowing departments to identify what's important to them, and to assess whether or not they're communicating and teaching these pillars of their thought. As mentioned before, post-secondary education is more complex and multifaceted than ever before. With so many sources of funding, prestige, and quality of education on the line, a program assessment must prove to be equally comprehensive and ever-changing. Although an assessment of student knowledge of course content is valuable in order to improve the delivery of that content, it is but one piece of the whole.

References

- Ackerman, P. L., & Kanfer, R. (2009). Test length and cognitive fatigue: an empirical examination of effects on performance and test-taker reactions. *Journal of Experimental Psychology: Applied*, 15(2), 163.
- American Psychological Association. (2016). Guidelines for the undergraduate psychology major: Version 2.0. *The American psychologist*, 71(2), 102.
- American Psychological Association. (2014). Strengthening the common core of the introductory psychology course. *Washington, DC: Author*.
- Association of American Colleges and Universities. (2009). Learning and assessment: Trends in undergraduate education.
- Association of American Colleges and Universities. (2016) Trends in learning outcomes assessment.
- Board of Educational Affairs Steering Committee. (2011). Principles for quality undergraduate education in psychology. *The American psychologist*, 66(9), 850.
- Campbell, D. T. (1960). Recommendations for APA test standards regarding construct, trait, or discriminant validity. *American Psychologist*, 15(8), 546.
- Campbell, J.P., McCloy, R.A., Oppler, S.H., & Sager, C.E. (1993). A theory of performance. In N. Schmitt & W. Borman (Eds.), *Personnel selection in organizations* (pp. 35-70). San

Francisco, CA: Jossey-Bass Publishers

Dunn, D. S., McCarthy, M. A., Baker, S., Halonen, J. S., & Hill IV, G. W. (2007). Quality

benchmarks in undergraduate psychology programs. *American Psychologist*, *62*(7), 650.

Ekuma, K. J. (2012). The importance of predictive and Face validity in employee selection and

ways of maximizing them: An assessment of three selection methods. *International Journal of Business and Management*, *7*(22), 115.

Griggs, R. A., Bujak-Johnson, A., & Proctor, D. L. (2004). Using common core vocabulary in text selection and teaching the introductory course. *Teaching of Psychology*, *31*(4), 265-269.

Haynes, S. N., Richard, D., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological assessment*, *7*(3), 238.

Halpern, D. F. et al (2010). *Principles for quality undergraduate education in psychology*.

Hatry, H., van Houten, T. Plantz, M. and Greenway, M. (1996). *Measuring Program Outcomes: A Practical Approach*. United Way of America, Alexandria, VA.

Homa, N., Hackathorn, J., Brown, C., Garczynski, A., Solomon, E., Tennial, R., Sanborn, U., & Gurung, R. A. R. (2013). How consistent are Introductory Psychology course syllabi? Examining objectives and content. *Teaching of Psychology*, *40*, 169–174.

Kraut, A. I., & Korman, A. K. (1999). *Evolving practices in human resource management: Responses to a changing world of work*. San Francisco, CA.: Jossey-Bass Publishers.

Landrum, R. E., Gurung, R. A., & Spann, N. (2012). Assessments of textbook usage and the relationship to student course performance. *College Teaching*, *60*(1), 17-24.

Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing research, 35*(6), 382-386.

Miller, B., & Gentile, B. F. (1998). Introductory course content and goals. *Teaching of Psychology, 25*(2), 89-96.

National Conference of State Legislatures (2015). Performance-based funding for higher education.

Sireci, S. G., & Geisinger, K. F. (1995). Using subject-matter experts to assess content representation: An MDS analysis. *Applied Psychological Measurement, 19*(3), 241-255.

Stoloff, M., McCarthy, M., Keller, L., Varfolomeeva, V., Lynch, J., Makara, K., Simmons, S., & Smiley, W. (2009). The undergraduate psychology major: An examination of structure and sequence. *Teaching of Psychology, 37*(1), 4-15.(2009).

Texas Tech University (2010). Assessing student learning in degree programs.

Appendix

A. Full Text of the Psychology Content Knowledge Assessment, With Answer Percentages

Q42

Dear Participant:

You are invited to participate in a research study regarding the assessment of student learning outcomes in the psychology department at this university. This study is being conducted by Dr. Andrea Lassiter and graduate student Colin Omori in the Industrial-Organizational Psychology program at Minnesota State University, Mankato. Your participation is voluntary. If you wish to participate in this study, please take 30-45 minutes to complete the test.

By completing and submitting the online questionnaire, you are providing us with valuable research data. If you choose not to participate, simply disregard this form. If you change your mind and wish to stop taking the survey, you may do so at any time by closing your web browser. Participation or nonparticipation will not impact your relationship with Minnesota State University, Mankato.

Your participation is confidential. Any identifying information will be deleted upon research completion by the researchers at Minnesota State. Furthermore, survey responses will be aggregated into larger groups and identifiable individual responses will never be made available to university staff. However, whenever one works with online technology there is always the risk of compromising privacy, confidentiality, and/or anonymity.

If you would like more information about the specific privacy and anonymity risks posed by online surveys, please contact the Minnesota State University, Mankato Information and Technology Services Help Desk (507-389-6654) and ask to speak to the Information Security Manager. The risks of participating are no more than are experienced in daily life. In addition to attaining one point of extra credit (which will be received by all who take this test), your responses may enable the university to better assess the knowledge of its students. Individuals with disabilities may obtain the questionnaire in an alternate format on request.

If you have any questions regarding this study or would like an additional copy of this letter, please contact Colin Omori at (507) 766-3357 or contact Dr. Andrea Lassiter at (507) 389-5734. For questions concerning for participants' rights and research-related injuries, contact Dr. Barry Ries at Minnesota State University, Mankato at (507) 389-1242. Selecting "Yes" below will indicate your informed consent to participate and indicate your assurance that you are at least 18 years of age. Selecting "No" below will indicate that you do not wish to participate in the

study. Please print a copy of this page for your future reference. Thank you for your participation.

Colin Omori

Industrial-Organizational Psychology Masters Candidate

Minnesota State University, Mankato

colin.omori@mnsu.edu

(507) 766-3357

IRBnet ID: 1051465

- Yes
- No

Condition: No Is Selected. Skip To: End of Survey.

Q44 What is your current status at this university?

- Freshman
- Sophomore
- Junior
- Senior

Q42 What is your current GPA?

- 3.6-4.0
- 3.2-3.599
- 2.8-3.199
- 2.4-2.799
- 2.0-2.399
- 1.6-1.999
- 1.2-1.599
- 0.8-1.199
- 0.4-.799
- 0.0-.399

Q43 How many Psychology classes have you taken while in college or through a program such as Advanced Placement (AP) or Post-Secondary Enrollment Options (PSEO)?

- 0
- 1-2
- 3-4
- 5-6
- 7-8
- 9+

1 Which of the following is a defining difference between major depressive disorder and bipolar disorder?

- Major depressive disorder involves only depressive symptoms, but bipolar disorder involves both manic and depressive symptoms. (N = 77, 85.56%)**
- People with major depressive disorder suffer from intense anxiety, but those with bipolar disorder do not. (N = 3, 3.33%)
- Bipolar disorder can be treated with medications, but major depressive disorder cannot. (N = 1, 1.11%)
- Psychotic features can occur in bipolar disorder but not in major depressive disorder. (N = 7, 7.78%)
- Genetic causes have been established for bipolar disorder but not for major depressive disorder. (N = 2, 2.22%)

2 Which of the following is characterized by a long history of dramatic complaints about faked medical conditions?

- Factitious disorder (N = 59, 65.56%)**
- Somatoform disorder (N = 19, 21.11%)
- Anxiety disorder (N = 5, 5.56%)
- Mood disorder (N = 4, 4.44%)
- Adjustment disorder (N = 3, 3.33%)

3 Which of the following is an example of recognition testing?

- “Is this word one of the words I showed you earlier?” (N = 65, 72.22%)**
- “What did you have for breakfast yesterday?” (N = 3, 3.33%)
- “What is the state capital of Massachusetts?” (N = 2, 2.22%)
- “Can you list all of the bones in the human body, in any order you choose?” (N = 4, 4.44%)
- “On the list I showed you earlier, which words began with the letter ‘m’?” (N = 16, 17.78%)

4 Autistic disorder is a developmental condition characterized by all of the following EXCEPT

- intolerance of change (N = 8, 8.89%)
- communication problems (N = 4, 4.44%)
- ritualistic repetitive behavior (N = 13, 14.44%)
- weak attachment to others (N = 28, 31.11%)
- hyperactivity (N = 37, 41.11%)**

5 Which of the following individuals is at the highest risk of developing schizophrenia?

- Mary, whose identical twin sister has schizophrenia (N = 52, 57.78%)**
- Edward, whose fraternal twin sister has schizophrenia (N = 3, 3.33%)
- Jamal, whose mother has schizophrenia (N = 15, 16.67%)
- Peter, whose father has schizophrenia (N = 9, 10%)
- Saul, whose maternal grandmother and maternal grandfather have schizophrenia (N = 11, 12.22%)

6 Which of the following individuals is most likely to benefit from taking a medication that is a selective serotonin reuptake inhibitor (SSRI)?

- Jack, who has lost his sense of identity and has wandered from his home to a distant city (N = 2, 2.22%)
- Tyler, who believes that he has several distinct personalities (N = 7, 7.78%)
- Andrea, who hears imaginary voices telling her that she will suffer a fatal accident (N = 3, 3.33%)
- Shannon, who feels helpless, worthless, and apathetic and who thinks that her life is without meaning (N = 70, 77.78%)**
- David, who falsely believes that other people are trying to kill him (N = 8, 8.89%)

7 Which of the following is an example of a delusion?

- Seeing a vision of an object that is not physically present (N = 27, 30%)
- Being erroneously convinced that one's neighbors are plotting to take over the world (N = 56, 62.22%)**
- Having to wash one's hands thoroughly every time one touches something (N = 2, 2.22%)
- Feeling pain in one's back even though there is no physical sign of injury (N = 5, 5.56%)
- Experiencing great anxiety when looking at a dangerous object (N = 0, 0.00%)

8 In an experiment, a pigeon was rewarded every twenty seconds regardless of the pigeon's behavior. By the end of the experiment, the frequency of the pigeon's grooming responses had significantly increased. Which of the following phenomena describes this result?

- Generalization (N = 19, 21.11%)
- Superstitious behavior (N = 11, 12.22%)**
- Discrimination (N = 0, 0.00%)
- Classical conditioning (N = 58, 64.44%)
- Modeling (N = 2, 2.22%)

9 Eric displays no regard for the rights of others, shows no remorse when he abuses others, and has a repeated pattern of stealing. It is most likely that he has which of the following personality disorders?

- Antisocial (N = 48, 53.33%)**
- Paranoid (N = 2, 2.22%)
- Narcissistic (N = 31, 34.44%)
- Histrionic (N = 4, 4.44%)
- Schizoid (N = 5, 5.56%)

10 One of the fundamental assumptions of the behaviorist approach is that

- the probability of a behavior depends on past consequences of that behavior (N = 48, 53.33%)**
- the probability of a behavior depends on people's beliefs and expectations (N = 18, 20.00%)
- unconscious thoughts influence one type of behavior and conscious thoughts influence another (N = 18, 20.00%)
- people have a free will that differentiates their behavior from that of other species (N = 5, 5.56%)
- the best way to uncover the reasons behind people's behavior is simply to ask them for self-reports (N = 1, 1.11%)

11 An experimenter who incorrectly rejects the null hypothesis commits which of the following kinds of error?

- A systematic error (N = 20, 22.22%)
- A random error (N = 4, 4.44%)
- A Type I error (N = 46, 51.11%)**
- A false alarm (N = 2, 2.22%)
- The fundamental attribution error (N = 18, 20.00%)

12 A 40-item vocabulary test was administered to a group of students. A second, similar test of vocabulary terms was administered to this same group of students approximately one week later. The researcher reported that the correlation between these two tests was $r = .90$. Which of the following can be stated regarding the two tests?

- They have strong reliability. (N = 64, 71.11%)**
- They have weak reliability. (N = 6, 6.67%)
- They have strong divergent validity. (N = 6, 6.67%)
- They have weak content validity. (N = 6, 6.67%)
- They have strong face validity. (N = 8, 8.89%)

13 On an attitude scale, the consecutively ordered scale points 1 through 7 represent different degrees of attitude strength, but the distance between any two adjacent scale points remains undetermined. Given what is known, this scale can be said to have the properties of

- a ratio scale (N = 17, 18.89%)
- a nominal scale (N = 26, 28.89%)
- a multidimensional scale (N = 7, 7.78%)
- an ordinal scale (N = 14, 15.56%)**
- an interval scale (N = 26, 28.89%)

14 A correlation of .20 is found between education and income in a representative sample of residents in city X. However, the correlation is not statistically significant. On the basis of this outcome alone, which of the following conclusions is most warranted?

- Increased education generally results in no increase in income in most cities. (N = 1, 1.11%)
- People with more education generally tend to earn more income in city X, but we do not know which is the cause and which is the effect. (N = 26, 28.89%)
- In city X, most of the people with advanced degrees are wealthy. (N = 12, 13.33%)
- In city X, the relationship between education and income is probably curvilinear. (N = 9, 10.00%)
- No significant relation between education and income was established. (N = 42, 46.67%)**

15 On a test with a positively skewed distribution, one student received the mean score, one student received the median score, and one student received the mode score. Which of these scores has the greatest percentile ranking?

- The mean (N = 18, 20.00%)**
- The median (N = 15, 16.67%)
- The mode (N = 19, 21.11%)
- All three scores are equal. (N = 4, 4.44%)
- It cannot be determined from the data given. (N = 34, 37.78%)

16 A marketing company rewards associates with a bonus for every 15 surveys they complete. Which of the following schedules of reinforcement best describes this pay arrangement?

- Fixed interval (N = 43, 47.78%)
- Fixed ratio (N = 32, 35.56%)**
- Variable interval (N = 7, 7.78%)
- Variable ratio (N = 4, 4.44%)
- Continuous (N = 4, 4.44%)

17 Jack has an irresistible urge to check his door lock exactly five times before he goes to bed. Although he realizes that his behavior is excessive and irrational, he feels quite anxious unless he performs this task in exactly the same manner each night. Jack's behavior is an example of

- an obsession (N = 16, 17.78%)
- a compulsion (N = 67, 74.44%)**
- a delusion (N = 5, 5.56%)
- an overvalued idea (N = 2, 2.22%)
- magical thinking (N = 0, 0.00%)

18 Psychologists studying sensation and perception are interested in research on phantom limb pain because such research

- shows that the gate-control theory of pain perception cannot be correct (N = 5, 5.56%)
- offers direct support for opponent-process theory (N = 10, 11.11%)
- demonstrates that pain signals must travel through the superior colliculus (N = 11, 12.22%)
- points out that pain can be perceived even when pain receptors are not being stimulated (N = 61, 67.78%)**
- illustrates that thresholds are typically higher for passive touch than for active touch (N = 3, 3.33%)

19 Research suggests that young girls who frequently watch television are more likely than other young girls to be highly gender stereotyped. We cannot conclude that television viewing leads to increased gender stereotyping in girls because

- the study is an example of an illusory correlation (N = 13, 14.44%)
- the study did not include adult women as well as young girls (N = 14, 15.56%)
- the study was probably based on a random sample of girls (N = 9, 10.00%)
- cause-and-effect conclusions cannot be drawn from these data (N = 52, 57.78%)**
- experimenter bias explains the relationship (N = 2, 2.22%)

20 A diagnostic test for a disease has 95 percent accuracy, but the incidence of the disease is low, only 1 in 1,000. This means that when this test is used for diagnosis, which of the following will be true?

- There will be many more misses than false alarms. (N = 12, 13.33%)
- There will be many more false alarms than misses. (N = 25, 27.78%)
- The number of false alarms and misses will be equal. (N = 10, 11.11%)
- If patients are told that they have the disease, there is a 95% chance that they actually have it. (N = 31, 34.44%)**
- If patients are told that they do not have the disease, there is a 95% chance that they do not actually have it. (N = 12, 13.33%)

21 In a memory experiment, participants study two lists of 15 unrelated words. Participants in the experimental group study list 1 and then study list 2, each for 3 minutes. Participants in the control group first look at a set of irrelevant pictures for three minutes and then study list 2 for three minutes. Which of the following results would be expected?

- The control group will recall all of the items in list 2. (N = 6, 6.67%)
- The control group will recall more items from list 2 than will the experimental group. (N = 46, 51.11%)**
- The experimental group will recall more items from list 2 than will the control group. (N = 20, 22.22%)
- The experimental group will recall approximately equal numbers of items from both list 1 and list 2. (N = 10, 11.11%)
- The experimental group will recall more items from list 1 than from list 2. (N = 8, 8.89%)

22 The most appropriate way for researchers to assess participants' implicit memory of a short essay would be to ask them to

- answer a series of multiple-choice questions about the essay (N = 28, 31.11%)
- remember as many specific sentences as possible from the essay (N = 11, 12.22%)
- summarize the gist of the essay, being as accurate as possible (N = 26, 28.89%)
- read the essay again, and see whether the participants read it faster than a comparable, but unfamiliar, essay (N = 10, 11.11%)**
- read the essay again, using a deeper level of processing; then measure the increase in recall with deeper processing (N = 15, 16.67%)

23 On Tuesday, Sam had fish for dinner. On Wednesday, he did not think about Tuesday's dinner at all. On Thursday, he thought about how good the fish had tasted on Tuesday. On Wednesday, therefore, Sam's memory of Tuesday's dinner was represented in

- short-term memory (N = 19, 21.11%)
- episodic memory (N = 42, 46.67%)**
- procedural memory (N = 3, 3.33%)
- sensory memory (N = 12, 13.33%)
- semantic memory (N = 14, 15.56%)

24 Neurotransmitters are released from what neuronal structure?

- Dendrite (N = 36, 40.45%)
- Myelin sheath (N = 5, 5.62%)
- Mitochondrion (N = 5, 5.62%)
- Soma (N = 6, 6.74%)
- Axon terminal (N = 37, 41.57%)**

25 People from collectivist cultures are more likely than people from individualistic cultures to engage in which of the following behaviors?

- Focusing on internal self-attributes (N = 6, 6.67%)
- Developing a large circle of friends (N = 38, 42.22%)
- Defining the self in relation to others (N = 37, 41.11%)**
- Being clear and direct in self-expression (N = 4, 4.44%)
- Showing a self-enhancing bias (N = 5, 5.56%)

26 How might the response tendency referred to as “social desirability” relate to an individual’s performance on a personality inventory?

- An individual may be more likely to respond randomly. (N = 7, 7.78%)
- An individual’s score will be high because of a high level of interest and involvement. (N = 10, 11.11%)
- An individual’s score will be adversely influenced by the individual’s interaction with the person administering the inventory. (N = 14, 15.56%)
- An individual may try to answer the questions in a way that places the individual in a more favorable light. (N = 56, 62.22%)**
- An individual may provide incorrect answers or misleading information that places the individual in a less favorable light. (N = 3, 3.33%)

27 Which of the following shows the application of social comparison theory?

- Janet feels poorly about herself because her teammate always beats her at tennis. (N = 58, 64.44%)**
- Bill likes Angela because she has the same hobbies that he has. (N = 16, 17.78%)
- Mark likes Freda because she is attractive. (N = 5, 5.56%)
- May reminds Frank of his old friend Marla. (N = 8, 8.89%)
- Howard pays extra attention to Charlene to get her to like him. (N = 3, 3.33%)

28 Which of the following is the best example of social exchange?

- “If you help me study for this exam, I will help you fix your car.” (N = 68, 76.40%)**
- “If I am in a good mood, you probably will be too.” (N = 7, 7.87%)
- “If I need to make a decision, I will ask you to do some research on the topic.” (N = 4, 4.49%)
- A person who lives in a small town is more likely to help others than a person who lives in a large city. (N = 4, 4.49%)
- A jury member shares her opinions with the other jury members. (N = 6, 6.74%)

29 Research regarding the risky shift phenomenon has indicated that decision making groups do not always shift in a more risky direction. The risky shift is now commonly seen as an example of which of the following?

- Social loafing (N = 11, 12.36%)
- Outgroup homogeneity (N = 19, 21.35%)
- Group polarization (N = 46, 51.69%)**
- Narrowing of attention (N = 7, 7.87%)
- The mere exposure effect (N = 6, 6.74%)

30 Which of the following concepts would create the most cognitive dissonance for a person with strong just-world beliefs?

- An uninvolved bystander (N = 13, 14.44%)
- A hard-working winner (N = 13, 14.44%)
- A convicted felon (N = 22, 24.44%)
- An innocent victim (N = 28, 31.11%)**
- A sore loser (N = 14, 15.56%)

31 While a board of directors was attempting to make a major decision, all members felt pressure to support the group's views and to reject ideas that were contrary to the opinions of the group. This example best illustrates

- group polarization (N = 21, 23.33%)
- group cohesiveness (N = 18, 20.00%)
- groupthink (N = 46, 51.11%)**
- social facilitation (N = 5, 5.56%)
- sunk costs (N = 0, 0.00%)

32 Kevin is 12 months old and has never seen a cat. When he first sees a cat, Kevin seems startled, but then he looks at his mother to see how she is responding. This is an example of which of the following?

- Accommodation (N = 4, 4.44%)
- Assimilation (N = 22, 24.44%)
- Prosocial behavior (N = 5, 5.56%)
- Self-efficacy (N = 1, 1.11%)
- Social referencing (N = 58, 64.44%)**

33 A preschool child says, "The sun is sad today." The child is demonstrating which of the following concepts?

- Animism (N = 50, 55.56%)**
- Centrism (N = 11, 12.22%)
- Conservation (N = 9, 10.00%)
- Irreversibility (N = 2, 2.22%)
- Object constancy (N = 18, 20.00%)

34 Which of the following best describes intellectual changes that typically occur in older people?

- Strategies for problem solving become more flexible and unconventional. (N = 14, 15.56%)
- Divergent thinking increases and convergent thinking decreases. (N = 22, 24.44%)
- Crystallized intelligence remains stable or increases slightly. (N = 20, 22.22%)**
- Fluid intelligence remains stable or increases slightly. (N = 4, 4.44%)
- Ability to memorize meaningful information decreases, but rote recall ability increases. (N = 30, 33.33%)

35 According to Erik Erikson's theory, the process of development occurs

- in three cognitive stages that are completed by five years of age (N = 16, 17.78%)
- throughout life in a series of psychosocial stages (N = 58, 64.44%)**
- as a consequence of biological maturation (N = 5, 5.56%)
- through increasing gains in moral reasoning (N = 5, 5.56%)
- in response to exposure to environmental stimuli (N = 6, 6.67%)

36 A factorial design is one that has

- one independent variable and one dependent variable (N = 28, 31.11%)
- one independent variable and two or more independent variables (N = 12, 13.33%)
- two or more independent variables and one dependent variable (N = 29, 32.22%)**
- two or more independent variables and no dependent variable (N = 16, 17.78%)
- no independent variable and two dependent variables (N = 5, 5.56%)

37 Which of the following is an example of confirmation bias?

- Janet thinks that there is a bus leaving at 3:00 PM, but she checks the schedule anyway to make sure this is so. (N = 30, 33.33%)
- Ahmet believes that his friend lied to him, so he reviews their recent conversation looking for instances in which the friend may have lied. (N = 38, 42.22%)**
- Philip is constantly checking to see if he locked the door to his house, left any lights on, or forgot to complete any of several common tasks. (N = 4, 4.44%)
- Although LaToya's travel agent told her to confirm that her flight was leaving on time, LaToya did not follow this advice, reasoning that it probably would leave on time. (N = 9, 10.00%)
- Roy has no confidence in his own ability to reason, so he readily accepts the conclusions of others. (N = 9, 10.00%)

38 Which statement is an example of a negative correlation?

- Delinquent children are likely to become adult criminal offenders. (N = 14, 15.73%)
- When the economy is good, more people go to college. (N = 5, 5.62%)
- When food consumption increases, weight increases. (N = 6, 6.74%)
- The more hours a teenager is employed, the lower his or her grades are. (N = 63, 70.79%)**
- Children who are rewarded for good behavior will grow up to lead productive lives. (N = 1, 1.12%)

39 Eighteen-year-old David received a large inheritance from his grandfather. Rather than using the money to pay for his tuition, he decided to buy a new sports car. According to Sigmund Freud, David shows signs of a

- strong id (N = 36, 40.00%)**
- strong ego (N = 31, 34.44%)
- strong superego (N = 17, 18.89%)
- weak preconscious (N = 5, 5.56%)
- weak collective unconscious (N = 1, 1.11%)

40 When he heard a child screaming, Phillip looked to see what other people were doing. Because they looked calm and unconcerned, Phillip decided that nothing needed to be done and did not go help the child. This is an example of

- the actor-observer effect (N = 37, 41.57%)
- the false-consensus effect (N = 16, 17.98%)
- pluralistic ignorance (N = 12, 13.48%)**
- social loafing (N = 17, 19.10%)
- cognitive dissonance (N = 7, 7.87%)