Article

SOCIETY FOR THE TEACHING OF PSYCHOLOGY

Teaching of Psychology
2020, Vol. 47(4) 274-284

DOI: 10.1177/0098628320945114
journals.sagepub.com/home/top
SAGE

Assessment of Scientific Inquiry and Critical Thinking: Measuring APA Goal 2 Student Learning Outcomes

Jon F. Mueller[1], Helen K. Taylor[2], Karen Brakke[3], Mike Drysdale[4], Kiesa Kelly[5], Gary M. Levine[6], and Jaclyn Ronquillo-Adachi[7]

Abstract

Goal 2 of the *APA Goals for Undergraduate Major in Psychology*, Scientific Inquiry and Critical Thinking, addresses the development of scientific reasoning and problem-solving, including effective research methods, in undergraduate psychology students. These skills serve as the foundation of not only introductory courses but also the entire psychology curriculum. In this article, we address why these skills are critical to help students interpret information and assess claims and review the current practice of measuring student reasoning in this domain. The reach of Goal 2 is broad, including the application of scientific reasoning to interpret psychological phenomena, the demonstration of psychology information literacy, the interpretation and execution of basic psychological research, and an understanding of sociocultural influences in scientific inquiry. We discuss the value of both formative and summative assessments in this area, offer examples of assessment tools currently in use, and provide a list of nationally normed assessments in critical thinking and scientific reasoning. We conclude with a discussion of limitations in current assessment practices and suggest possible future directions.

Keywords

assessment, student learning outcomes, scientific reasoning, critical thinking, information literacy, research design

Fake news. "Deepfakes."[1] Segmented media.
Distinguishing truth from fiction has always been challenging, but in today's crowded and confusing information world that task is even more formidable. What tools can a person employ to navigate this sea of information and misinformation, to examine claims and evaluate evidence to draw a reasonable conclusion? Certainly, it would be helpful if someone could approach this task as a scientist would, applying the rigorous modes of thinking that science has developed to answer difficult questions. Psychology students and, frankly all citizens, should develop the scientific reasoning and problem-solving skills necessary for thinking like a scientist, whether in the role of a psychologist, or a researcher, or anyone who encounters scientific claims in the media and must make judgments about their veracity and evaluate the quality of the evidence presented to support such claims.

Thus, it is not surprising that the American Psychological Association (APA), when formulating its set of goals for an undergraduate psychology major, included among its five primary domains a goal for scientific inquiry and critical thinking (APA, 2013, 2016). Even at the introductory level, the APA (2011) argues that students should be introduced to psychology as a scientific discipline. Although the APA Guidelines identify the importance of developing scientific reasoning and there is a multitude of classroom activities available with that goal in mind, assessment of these activities to determine if they are effective is less common (Stevens et al., 2016).

Psychology programs (or individual instructors) attempting to develop such skills will want to know whether those goals are being met. Yet, measurement of skill development, particularly higher order skill development, can be quite a challenging task. Thus, this article will examine the assessment of these skills—exploring tools to capture whether students have acquired and can demonstrate the capacity for scientific inquiry and critical thinking from a scientific perspective. We will share examples of good measures and discuss, to a lesser extent, processes for developing those skills.

[1] Department of Psychology, North Central College, Naperville, IL, USA
[2] Bellevue College, WA, USA
[3] Spelman College, Atlanta, GA, USA
[4] Vincennes University, IN, USA
[5] Tennessee State University, Nashville, TN, USA
[6] Edinboro University of Pennsylvania, PA, USA
[7] Cerritos College, Norwalk, CA, USA

Corresponding Author:
Jon F. Mueller, Department of Psychology, North Central College, 30 Brainard St., Naperville, IL 60540, USA.
Email: jfmueller@noctrl.edu

Goal 2 Outcomes and Indicators

Those who are unfamiliar with the discipline of psychology are often surprised by claims that psychology is a science. Goal 2, "Scientific Inquiry and Critical Thinking," underscores the importance of learning and exercising scientific underpinnings of the discipline.

"Scientific inquiry" encompasses the development of scientific reasoning and problem-solving, including being able to recognize, use, and evaluate research methods. Developing scientific inquiry skills includes learning active and deliberate approaches to investigating behavioral questions. In psychology, that means designing and implementing structured observations such as experiments, correlational studies, quasi-experiments, and surveys. Effective inquiry skills involve creative problem-solving, intentional explorations of existing literature, and heightened awareness of potential biases that might shape research pursuits.

"Critical thinking" is broadly defined by the ability to address skills in interpreting behavior as well as to apply and evaluate psychological theories that facilitate accurate interpretation. Critical thinking looks different depending on the context or discipline. However, one common element across different domains is that the critical thinker be able to interpret an argument or claim and evaluate the evidence given to support that claim. Students ultimately should be able to challenge a claim, whether it derives from a bona fide psychological theory or an "infomercial" by insisting on a review of pertinent evidence.

Students completing foundation-level courses (those that are typically completed in the first 2 years of the major or at the completion of a 2-year psychology program) should learn basic skills and concepts in interpreting behavior, studying research and evaluating its conclusions, and applying research design principles to drawing conclusions about psychological phenomena. Students completing a baccalaureate degree should focus on theory use as well as designing and executing research plans (APA, 2013). Table 1 displays specific outcomes and indicators for Goal 2 at the associate and baccalaureate levels.

Choosing Appropriate Goal 2 Assessments

Effective measurement of student learning outcomes requires congruence between the student learning outcome and the type of measurement tool used. Several dimensions can influence the choice of assessment, including the timing of the assessment (formative vs. summative assessment), the complexity of the cognitive skills to be assessed (objective testing vs. authentic assessment), and the origin of the assessment instrument (nationally normed vs. locally developed).

Assessment strategies can focus on skill development during instruction (e.g., formative) or on achievement at the end of the learning experience (e.g., summative). Formative assessments can be used to provide students with opportunities for practice of skills, feedback on that practice, and reflection upon performance and feedback. In contrast, instructors use summative assessment in exams, papers, and projects, so students can demonstrate how well they understand and apply what they have learned in a course or at the end of a program.

Strategies can also focus on traditional objective testing or authentic assessment strategies. According to Haladyna (1999), traditional objective measures tend to focus on the two lowest cognitive levels of Bloom's Taxonomy (knowledge and comprehension). In contrast, authentic assessment strategies strive to capture a student's ability to apply, analyze, evaluate, and synthesize understanding or skill via a more authentic form of assessment in which students "perform real-world tasks that demonstrate meaningful application of essential knowledge and skills" (Mueller, 2019, para. 1).

A variety of nationally normed tests can be deployed for summative purposes to measure Goal 2. Table 2 provides a roster of objective tests that can be used for those purposes. Sharing a common national assessment across programs is consistent with the call for consistency in content for the discipline much as other sciences have developed, with accompanying standardized tests that are widely used (Gurung et al., 2016). On the other hand, standardized assessments are typically prone to a greater teacher–test mismatch than locally created measures. That is, the content within a national test often does not align as well with the content taught in a particular institution or classroom compared to a locally created assessment (Popham, 1999).

Thus, given the value of tailoring an assessment to best capture the outcome(s) taught, the remainder of the article is dedicated to examples of locally developed, authentic assessments with links to specific Goal 2 outcomes. In some cases, the exemplars are showcased in the Summit on National Assessment in Psychology website (pass.apa.org). We offer the link to the website to access detailed instructions and rubrics. Some examples represent more recently emerging exemplars.

Authentic Goal 2 Assessments

Outcome 2.1: Use Scientific Reasoning to Interpret Psychological Phenomena

Mythbusting. Assignments that ask students to evaluate popular claims or myths using scientific reasoning have been found to be an effective way to build competence in this outcome area. For example, Blessing and Blessing (2010) assigned small groups of students to present credible research on whether their assigned psychological myth was supported or "busted." In terms of assessment, a homegrown critical thinking test was used as a pre- and postmeasure. Two sections of introductory psychology students participated in the experimental condition, while two other sections were taught without this intervention. In comparison to students who did not participate in this assignment, students in the experimental condition improved in their ability to engage in critical thinking when applied to psychological issues.

Table 1. Goal 2 Outcomes and Indicators.

Outcomes Students Will	Foundation Indicators Students Will	Baccalaureate Indicators Students Will
2.1 Use scientific reasoning to interpret psychological phenomena	2.1a Identify basic biological, psychological, and social components of psychological explanations (e.g., inferences, observations, operational definitions, interpretations)	2.1A Describe the value and limitation of using theories to explain behavioral phenomena
	2.1b Use psychology concepts to explain personal experiences and recognize the potential for flaws in behavioral explanations based on simplistic, personal theories	2.1B Develop plausible behavioral explanations that rely on scientific reasoning and evidence rather than anecdotes or pseudoscience
	2.1c Use an appropriate level of complexity to interpret behavior and mental processes	2.1C Incorporate several appropriate levels of complexity (e.g., cellular, individual, group/system, societal/cultural) to explain behavior
	2.1d Ask relevant questions to gather more information about behavioral claims	2.1D Generate alternative explanations based on perceived flaws in behavioral claims
	2.1e Describe common fallacies in thinking (e.g., confirmation bias, post hoc explanations, implying causation from correlation) that impair accurate conclusions and predictions	2.1E Use strategies to minimize committing common fallacies in thinking that impair accurate conclusions and predictions
2.2 Demonstrate psychology information literacy	2.2a Read and summarize general ideas and conclusions from psychological sources accurately	2.2A Read and summarize complex ideas accurately, including future directions, from psychological sources and research
	2.2b Describe what kinds of additional information beyond personal experience are acceptable in developing behavioral explanations (i.e., popular press reports vs. scientific findings)	2.2B Describe the characteristics and relative value of different information sources (e.g., primary vs. secondary, peer reviewed vs. nonreviewed, empirical vs. nonempirical)
	2.2c Identify and navigate psychology databases and other legitimate sources of psychology information	2.2C Develop a comprehensive strategy for locating and using relevant scholarship (e.g., databases, credible journals) to address psychological questions
	2.2d Articulate criteria for identifying objective sources of psychology information	2.2D Evaluate psychology information based on the reliability, validity, and generalizability of sources
	2.2e Interpret simple graphs and statistical findings	2.2E Interpret complex statistical findings and graphs in the context of their level of statistical significance, including the influence of effect size, and explain these findings using common language
2.3 Engage in innovative and integrative thinking and problem-solving	2.3a Recognize and describe well-defined problems	2.3A Describe problems operationally to study them empirically
	2.3b Apply simple problem-solving strategies to improve efficiency and effectiveness	2.3B Select and apply the optimal problem-solving strategy from multiple alternatives
	2.3c Describe the consequences of problem-solving attempts	2.3C Evaluate the effectiveness of selected problem-solving strategies
2.4 Interpret, design, and conduct basic psychological research	2.4a Describe research methods used by psychologists including their respective advantages and disadvantages	2.4A Evaluate the effectiveness of quantitative and qualitative research methods in addressing a research question
	2.4b Discuss the value of experimental design (i.e., controlled comparisons) in justifying cause–effect relationships	2.4B Limit cause–effect claims to research strategies that appropriately rule out alternative explanations
	2.4c Define and explain the purpose of key research concepts that characterize psychological research (e.g., hypothesis, operational definition)	2.4C Accurately identify key research concepts in existing and proposed research projects
	2.4d Replicate or design and conduct simple scientific studies (correlational or two-factor) to confirm a hypothesis based on operational definitions	2.4D Design and conduct complex studies to confirm a hypothesis based on operational definitions
	2.4e Explain why conclusions in psychological projects must be both reliable and valid	2.4E Design and adopt high-quality measurement strategies that enhance reliability and validity
	2.4f Explain why quantitative analysis is relevant for scientific problem-solving	2.4F Use quantitative and/or qualitative analyses to argue for or against a particular hypothesis

(continued)

Table 1. (continued)

Outcomes Students Will	Foundation Indicators Students Will	Baccalaureate Indicators Students Will
	2.4g Describe the fundamental principles of research design	2.4G Apply knowledge of research skills necessary to be an informed consumer of research or critic regarding unsupported claims about behavior
2.5 Incorporate sociocultural factors in scientific inquiry	2.5a Relate examples of how a researcher's value system, sociocultural characteristics, and historical context influence the development of scientific inquiry on psychological questions	2.5A Recognize the systemic influences of sociocultural, theoretical, and personal biases on the research enterprise and evaluate the effectiveness with which researchers address those influences in psychological research
	2.5b Analyze potential challenges related to sociocultural factors in a given research study	2.5B Design studies that effectively address the effects of sociocultural factors
	2.5c Describe how individual and sociocultural differences can influence the applicability/generalizability of research findings	2.5C Evaluate and design research with respect to controls for variations in behavior related to individual and sociocultural differences that can influence research outcomes
	2.5d Identify under what conditions research findings can be appropriately generalized	2.5D Evaluate the generalizability of specific findings based on parameters of the research design, including caution in extending Western constructs inappropriately

Source. APA (2013).

Infomercial critiques. Similarly, psychologists have utilized pseudoscience as a tool to foster critical thinking and information literacy (Adam & Manson, 2014; Lawson & Brown, 2018). For example, Adam and Manson designed a study to increase students' ability to think critically by having students watch and discuss infomercials. Students needed to identify flaws in infomercials that contained pseudoscientific claims. Using a pretest–posttest design, the researchers found that students were better able to critically evaluate claims after engaging in this activity than those exposed to a straight lecture on scientific reasoning.

Evaluating media claims. This article assignment asks students to locate a claim in the media and evaluate how well the evidence provides support for the claim. A link provides examples of scientific claims, specific instructions, and a grading rubric (http://pass.apa.org/docs/evaluating-a-claim-about-human-behavior-and-the-evidence-presented-to-support-it/).

Judging research evidence. This brief assignment provides students with a causal claim and seven varied pieces of evidence intended to support the claim, ranging from personal opinion to anecdotes to scientific research. Students judge how well each piece of evidence addresses the claim presented. The link provides a rubric to evaluate the quality of student responses (http://pass.apa.org/docs/evaluating-evidence-in-support-of-research-questions-2/).

Teaching students to critically evaluate claims and the evidence used to support them is not a simple task that can be completed in just a few lessons. In fact, numerous studies have discovered that even deliberate and repetitive practice and feedback on scientific reasoning skills may not produce students fluent at these tasks (e.g., Penningroth et al., 2007; Sibulkin & Butler, 2019). For example, to emphasize that correlation is not causation, Sibulkin and Butler taught their students to write "reverse causality" statements (i.e., if A and B are correlated, it is possible that A causes B, but it is also possible that B causes A). In analyzing test questions addressing reverse causality, they found that only 45% of student explanations were correct despite multiple instructional efforts and opportunities for practice.

Similarly, Mueller and Coon (2013) embedded formative assessments regularly throughout a 10-week term to provide practice and feedback for scientific thinking skills. Students made progress, but there was still considerable room for improvement. For example, at the beginning of the term, only 6.25% of students could correctly identify at least four of five definitional statements as correlational, while 42.8% of students made such an identification by the end of the term. As was also true of the Sibulkin and Butler (2019) study, students who did improve were the more capable students. Thus, students can learn to use scientific reasoning to interpret psychological phenomena, but considerable practice, feedback, and reflection will be required, particularly for weaker students.

For example, Bailey et al. (2019) recently investigated whether emphasizing positive dispositions such as self-efficacy and identity could further enhance the development of critical thinking skills in a research methods setting. Bartels and Griggs (2019) discuss how recent reexaminations of the Stanford Prison Experiment archives provides opportunities for students to apply scientific inquiry and critical thinking skills to classic studies. Alternatively, as described below, instructors may work backward from summative assessments to build formative assessments that systematically encourage the necessary practice, feedback, and reflection (e.g., Wentworth & Whitmarsh, 2017).

Table 2. Nationally Normed Assessments.

California Critical Thinking Skills Tests (CCTST) Insight Assessment http://www.insightassessment.com/About-Us/California-Critical-Thinking-Skills-Test-Family	The CCTST represents a family of tests for different populations from elementary school through doctoral levels. The tests target analytic skills and information interpretation from charts, texts, and images.
Cambridge Thinking Skills Assessment (TSA) Admissions Testing Service www.admissionstestingservice.org/our-services/thinking-skills/tsa-cambridge/about-tsa-cambridge	The Cambridge TSA is a multiple-choice test involving 25 questions measuring problem-solving and 25 questions examining critical thinking. The test has been used since 2001 to assess critical thinking outcomes.
Collegiate Assessment of Academic Proficiency Critical Thinking Test (CAAP) www.act.org/caap	The CAAP assessment program enables colleges and universities to assess and evaluate both general education program outcomes and outcomes from institutional student learning outcomes.
Collegiate Learning Assessment (CLA) Council for Aid to Education www.collegiatelearningassessment.org	The CLA measures critical thinking, analytic reasoning, problem-solving, and written communication skills using open-ended questions to determine value-added to student learning from college programs.
Cornell Critical Thinking Test (CCTT) Ennis & Millman https://www.criticalthinking.com/cornell-critical-thinking-test-level-z.html	The CCTT (Level Z) presents a series of scenarios and presents an appropriate and logical solution from multiple options. The test evaluates deduction, induction, credibility, and identification of assumptions.
Ennis–Weir Critical Thinking Essay Test Ennis & Weir https://www.academia.edu/1847582/The_Ennis-Weir_Critical_Thinking_Essay_Test_An_Instrument_for_Teaching_and_Testing	The Ennis–Weir Test requires the production of an essay prepared as a letter to the editor of a newspaper. To measure critical thinking ability, the test focuses on the identification of main point, reasons, and assumptions as well as seeing other possible explanations.
Halpern Critical Thinking Assessment (HCTA) Halpern https://www.schuhfried.com/test/HCTA	The HCTA test includes five dimensions of critical thinking—verbal reasoning, argument analysis, thinking as hypothesis testing, likelihood and uncertainty, and decision making and problem-solving—using both constructed and recalled answers in response to 25 scenarios.
International Critical Thinking Essay Test Paul and Elder www.criticalthinking.org/pages/international-critical-thinking-test/619	This essay-based exam involves two parts—analysis and assessment—of writing prompts and is graded holistically by individual graders.
Measure of Academic Proficiency and Progress (MAPP) Educational Testing Service https://www.ets.org/s/mapp/pdf/5018.pdf	The MAPP evaluates general education skills, including critical thinking, reading, writing, and mathematics. The test has been included in the Degree Qualifications Profile and has also been supported by the Lumina Foundation.
Proficiency Profile Educational Testing Service www.ets.org/proficiencyprofile/about	The Proficiency Profile provides institutional data about program value by examining critical thinking, writing, reading, and mathematics and has been selected by the Voluntary System of Accountability (VSA) as a primary measure of general education quality.
Psychological Critical Thinking Exam (updated) Lawson https://journals.sagepub.com/doi/10.1177/0098628315587624	This test provides a way to differentiate gains in critical thinking related to psychology.
Test of Psychological Knowledge and Misconceptions Bensley and Lilienfeld (2015)	This test is designed to identify psychological misconceptions as well as confidence ratings of the misconception knowledge.
Watson–Glaser Critical Thinking Appraisal Watson & Glaser https://us.talentlens.com/store/ustalentlens/en/Store/Ability/Watson-Glaser-Critical-Thinking-Appraisal-III/p/100001976.html	The approach in this test involves recognizing assumptions, evaluating arguments, and drawing conclusions and is used in academic and workplace settings to identify talent.
Medical College Admission Test (MCAT) (section on psychological, social, and biological foundations of behavior) Association of American Medical Colleges.	More than half the questions in this section are based on introductory psychology, with many addressing application of scientific reasoning.

Outcome 2.2: Demonstrate Psychology Information Literacy

Hands-on classroom practice. Psychology course assignments that ask students to evaluate published research studies offer a promising avenue for increasing psychology information literacy. Stevens and Witkow (2014) asked students to interpret and evaluate a research study, including a data set and table, as an in-class small-group instructional module. In follow-up studies, Stevens et al. (2016) expanded the intervention to eight in-class scientific reasoning modules offered in both a baccalaureate liberal arts college and a community college. Aligned with the APA's undergraduate guidelines and scientific reasoning outcomes (2013) as well as other standards, each module occupied 30–45 min of class time and addressed a topic in introductory psychology using an empirical article and lesson plan. The instructor assessed students using a pre- and postscientific reasoning questionnaire based on a combination of a Medical College Admission Test (MCAT) passage set and questions and the researchers' own questions. Two different studies provided strong evidence that specific targeted activities can strengthen scientific reasoning in introductory psychology classes.

The misconceptions test. Based on work by Lilienfeld et al. (2010), this test asks students to identify the strength of their beliefs regarding 20 common misconceptions in psychology. As a pre–post measure, this test can be given the first day of class and at the end of the term to assess gains in scientific reasoning. It can also be used at the program level to compare reasoning in students about to graduate from that of beginning psychology students (http://pass.apa.org/docs/the-misconceptions-test/).

Thanks for the memory. This assignment has three different activities designed to assess the understanding of long-term memory. In the first two activities, students define terms related to long-term memory (episodic, semantic, etc.) and identify the conceptual relationships among the terms. In the third activity, students then write a short paper and apply the terms to an event in their personal lives. This activity addresses multiple goals including the demonstration of psychology information literacy (http://pass.apa.org/docs/thanks-for-the-memory-exploring-the-multistore-memory-model/).

The superstitious pigeon. This assessment requires students to recognize elements of conditioning and provides practices in summarizing and paraphrasing disciplinary information. Students receive an excerpt from a journal article by B. F. Skinner and draw conclusions using quotes from the article as supporting evidence, thus gaining practice in summarizing psychological information and demonstrating psychology information literacy (http://pass.apa.org/docs/the-superstitious-pigeon-summarizing-interpreting-and-paraphrasing-psychological-literature/).

Information literacy. Compiled by Jon Mueller, this website provides dozens of links to general measures of information literacy that are freely available online. Critical aspects of information literacy, such as the abilities to define a research question, locate appropriate sources, access relevant information from those sources, and evaluate the accuracy, authority, currency, and relevancy of the information, can be measured through a variety of forced choice or authentic assessments linked to on this page or by applying one of the many rubrics found here (http://jfmueller.faculty.noctrl.edu/infolitassessments.htm).

Distinguishing correlational from experimental research. In this brief activity, instructors provide five examples of research. For each study students identify if the study is correlational or experimental. If correlational, they draw a scatter plot and predict the correlational coefficient (r). If experimental, they draw a bar graph and label axes. As a possible extension, students can explain what conclusions the researchers can make regarding the relationship between the variables (i.e., causal inference or correlational). This activity assesses psychological information literacy along with student understanding of key distinctions in psychological research (http://pass.apa.org/docs/distinguishing-correlational-vs-experimental-research/).

Distinguishing correlational claims and research from causal ones. This web page provides more than 100 headlines found in the popular media that make either correlational or causal claims. Several formative activities designed by various instructors are also listed which suggest a number of ways faculty can use these headlines to teach scientific thinking skills. Furthermore, each headline is linked to the popular press article so that students can learn to evaluate whether a headline's claim is supported by the research in the article (http://jfmueller.faculty.noctrl.edu/100/correlation_or_causation.htm).

Locating and evaluating sources and creating an annotated bibliography. This formative assessment activity can be modified for foundational- or baccalaureate-level students. At the foundational level, students are asked to complete several tasks involving locating and evaluating database articles, for example, comparing the results of a study to their own experience and/or locating and comparing a journal article to a piece from the popular press on the same topic. At the baccalaureate level, students write a hypothesis, plan a search, locate 10 peer-reviewed articles, construct an annotated bibliography, and evaluate each article. Specific instructions and a scoring rubric are included (http://pass.apa.org/docs/locating-and-evaluating-sources-and-creating-an-annotated-bibliography/).

Progressive research evaluation journal assignment. This journal assignment comes from a high school chemistry unit on scientific literacy but includes thoughtful assignments that address interpreting research (Outcome 2.4) and information literacy. Students keep a weekly journal with specific instructions to evaluate sources of information, starting with newspaper articles and moving to research studies. By the end of the 6-week assignment, students can be assessed on identifying scientific principles, noting the author's perspective, identifying variables, and evaluating support for the author's conclusion (http://jfmueller.faculty.noctrl.edu/toolbox/examples/hayner03/journalentries.pdf).

The rubric can be found here (http://jfmueller.faculty.noctrl.edu/toolbox/examples/hayner03/journalrubric.pdf).

Outcome 2.3: Engage in Innovative and Integrative Thinking and Problem-Solving

Problem-based scenarios. Using a problem-based learning approach promises another effective vehicle for strengthening critical thinking skills. While commonly used in other disciplines, such as medicine and business, a course focus on problem-solving is less common in psychology but may provide an effective way to engage students as well as a means of assessing outcomes. Muehlenkamp et al. (2015) designed an introductory psychology course at the University of Wisconsin–Eau Claire around four problem scenarios provided by the instructor (and an additional problem identified by students), each linked to several chapters of content. For each unit/problem scenario, students worked in groups analyzing related empirical articles and teaching each other key concept/content areas related to the problems. A combination of questions from standardized assessment tools measured four components of critical thinking: use of higher level thinking, critical thinking processes, tools for critical thinking, and analysis of psychological research. Pre- and postquestionnaires found that students engaged in the problem-solving approach reported significant improvements in critical thinking compared to traditionally taught classes.

A problem-based learning approach in social psychology. To encourage students to learn how to apply psychological theories and concepts to address real-world problems, this project uses a problem-based learning method by introducing a "messy" problem before students have acquired the knowledge to solve it. Specifically, students must develop strategies to improve recycling by students in the dorms and at the college's primary restaurant on campus. Students direct the learning by (a) defining the problem, (b) identifying what they already knew, (c) identifying what they still needed to do, (d) determining how they would find the needed information, repeating this process until sufficient information had been collected, (e) generating possible solutions, (f) evaluating the solutions, and finally (g) selecting one or more solutions to the problem to be implemented. A description of the project's solution paper, rationale paper, and journal reflections can be found at the following site (http://jfmueller.faculty.noctrl.edu/240/assign.htm).

Testing proverbs. Another problem-based learning project emphasizes the acquisition of study design skills. Working in small groups, students take well-known falsifiable proverbs, develop hypotheses to test them, recruit and assess participants, analyze the data, and present their findings in a poster format. This approach allows students to develop key skills to systematically experiment with an idea of using proverbs as a proxy for theory (http://eprints.bournemouth.ac.uk/31130/).

Outcome 2.4: Interpret, Design, and Conduct Psychological Research

Regardless of whether students are taking a psychology class at the foundation or baccalaureate level, they should be able to demonstrate an understanding of the fundamental principles of research design and be able to replicate, design, and ideally conduct at least simple scientific studies. One of the primary ways to assess competence in this area is through course assignments in which students analyze and interpret studies or describe a study they design themselves. For example, Marek et al. (2004) required sophomore-level research methods' students to participate in the design, data collection, and analyses of a prescribed theme using a preselected set of materials. The quality of the final reports along with student self-assessments suggested that a hands-on research project is an excellent approach to teaching and assessing understanding of the nuts and bolts of psychological research. Even at the introductory level, coming up with a simple design to test a hypothesis provides a way to understand and apply basic research design.

Assessing statistical literacy. In this assignment, students enter, analyze, interpret, and write up data based on the study by Loftus and Palmer (1974) investigating the role of language on eyewitness testimony. In a recent replication of this study, college students estimated the speed of cars either "hitting" or "smashing" into each other after seeing a video clip of a car accident. For the assignment, students identify and set up the variables in a statistical software package, enter the data, create a graph from the data, and interpret and write up the results (http://pass.apa.org/docs/assessing-statistical-literacy/).

Conducting a study and writing up a research report. In response to a list of sample topics, students select one of the classic theories in developmental psychology and design, conduct, and analyze results from a study that tests that theory. Students must conduct a literature review, collect data, and write an APA-style research report based on the project. This substantial assignment takes students through each step of conducting and writing up research and can be adapted to other fields of psychology (http://pass.apa.org/docs/conducting-a-study-and-writing-up-a-research-report-2/).

Designing and reporting a simple experimental study. Designed for foundation level (introductory) classes, this assignment asks students to choose one of eight social psychology hypotheses and design an experiment to test the hypothesis. The assignment asks them to identify the purpose, hypothesis, method, subject selection and assignment, and conditions and variables of the study. Students speculate on the potential results of the experiment but do not actually collect data (http://pass.apa.org/docs/designing-and-reporting-a-simple-experimental-study/).

Finding flaws in claims: Multiple choice assessment. This assignment provides four "press releases," each followed by a short series of multiple-choice questions testing understanding of research terms, methods, and flaws. This formative assessment can provide a quick measure of understanding and can be used at different levels, most aptly in a research methods course. It can be designed as a paper/pencil measure or using online survey software and can be adapted for other courses (http://pass.apa.org/docs/finding-flaws-in-claims-multiple-choice-assessment/).

The effect of music on performance. This formative assessment uses an in-class experiment in which nine randomly selected student "subjects" complete a simple task while music of various tempos is played in the background. After identifying a hypothesis, the rest of the class observes the effect of no music, slow music, and fast music conditions on completion of the task. Following the trials, the classes discusses and/or writes responses to questions regarding identification of variables, results, possible confounding variables, and the impact of music in their own lives (http://pass.apa.org/docs/the-effect-of-music-on-performance/).

Random assignment: It's all in the cards. To illustrate the concept of random assignment, this demonstration shows that groups are more likely to be "statistically equivalent" when created by chance. Following a demonstration of an experiment (one on levels of processing is provided in this link), the instructor discusses confounding variables and how random assignment can eliminate these potential confounds. The instructor then passes around one playing card per student from a deck of cards and asks students, for example, to stand if they have a red card. From this group, students will stand or sit in response to questions about variables like gender, height, birth order, athleticism, sleep deprivation, and so on, demonstrating a random distribution of variables. Students then write an explanation of what might be a confounding variable in a list of experiments, and "how" and "why" random assignment is important in each of those studies (http://pass.apa.org/docs/random-assignment-its-all-in-the-cards/).

Outcome 2.5: Incorporate Sociocultural Factors in Scientific Inquiry

Evaluating sociocultural factors in scientific inquiry. Students should be able to recognize the importance of the cultural and historical context of scientific inquiry as well as attend to potential challenges of sociocultural differences when designing and interpreting findings. Designing studies that effectively address the influences of cultural differences is an important challenge at the baccalaureate level, whereas the ability to identify the appropriateness of generalizing findings can be assessed at any level of psychology instruction. Recent attention to bias in research and the promotion of culturally appropriate research methods (APA, 2017) can be a focus of class discussion, followed by an assignment asking students to seek out bias and evaluate the generalizability of previous studies.

One such assignment designed by Ronquillo (http://pass.apa.org/docs/evaluating-sociocultural-factors-in-scientific-inquiry/) asks students to read a classic study, specifically the Bandura et al. (1961) study on imitation of aggressive models. After reading the article and a short summary, students respond to questions about how the researchers' values and social–cultural backgrounds could have influenced the research question and how it was tested as well as how the results may have been affected by sociocultural factors.

Analyzing potential sociocultural challenges in scientific inquiry. In this assignment, students read a study on the effects of social inclusion on self-regulation by Baumeister et al. (2005). Students then identify how the researchers' values and/or cultural background might have influenced the research question and how it was tested, consider alternative explanations of the results, and redesign the study with sociocultural factors in mind (http://pass.apa.org/docs/analyzing-potential-sociocultural-challenges-in-scientific-inquiry/).

A Final Word on the Value of Formative Assessments for Skill Development

As mentioned above, formative assessments can not only can provide a check for student understanding but also foster the learning of the concepts or skills being assessed. Good skill development generally requires four components:

- Instruction/modeling of the skill
- Practice of the skill
- Feedback on the practice
- Reflection on the performance and the feedback on it

Formative assessments are excellent tools to promote the last three steps—practice, feedback, and reflection—once instruction and/or modeling has introduced the concept or skill. Through careful use of formative assessments, a variety of tasks can serve as opportunities for practicing what has been taught, for receiving or giving feedback on the practice, and for meaningful reflection on the practice itself as well as the feedback given. Thus, formative assessments can be valuable tools in the development of student learning, especially when it comes to scientific reasoning.

An example used effectively by one of the authors (J.M.) is to give frequent assignments asking students to read brief research articles or claims in the media and then respond to several questions in writing, followed by a class discussion (http://jfmueller.faculty.noctrl.edu/100/printsched.htm). Students evaluate the headlines and articles linked to this author's website on correlation and causation (http://jfmueller.faculty.noctrl.edu/100/correlation_or_causation.htm). In both cases, students gain practice in distinguishing between descriptive, correlational, and causal methods as well as in evaluating evidence for media claims.

Frequent, low-threat assignments like these can be effective at engaging students in scientific thinking practice, feedback, and reflection. Students are not graded for accuracy but whether or not they put in a "good faith effort," a grading technique developed by Walvoord and Pool (1998). This approach not only encourages student participation but is relatively easy to grade. Even weaker and less motivated students completed the assignments and benefited from class discussion and instructor feedback. Frequent practice facilitated strong performance on summative assessments at the end of the term.

Limitations and Future Directions

As evidenced by the breadth and depth of the example strategies above, the enterprise of assessing Goal 2, Scientific Inquiry and Critical Thinking, is alive and well in psychology classrooms at both the foundational and baccalaureate level. It is fair to say that, of all of the APA goals for the undergraduate major in psychology, the application of scientific reasoning to interpret psychological phenomena provides the foundation of every psychology course. We have developed a multitude of tools to measure psychology information literacy, the interpretation and execution of basic psychological research, and an understanding of sociocultural influences in scientific inquiry. Psychology instructors have a large toolbox of assessment techniques at their disposal, ranging from nationally normed tests to brief in-class activities to in-depth assignments, most of which include scoring rubrics and the potential for pre-post measurement.

A recent study found that the most commonly used assessment tools are reported to be rubrics, homegrown assessment measures, local exams, and final projects (Norcross et al., 2016). Both formative and summative measures provide students the opportunity to test themselves in their understanding of scientific reasoning and methods and can be used in the individual classroom as well as for the departmental assessment of graduating seniors.

While the assessment tools described above offer exemplary ways to measure the development of scientific reasoning, many educators argue that there is still a need for a nationally consistent and coordinated assessment process tied to the Guidelines, particularly for introductory psychology students (Gurung et al., 2016). As identified in a recent study of undergraduate psychology programs, although the majority of institutions participate in regular program reviews and do conduct some form of course assessment, more standardized assessment measures are called for (Norcross et al., 2016). However, as noted above, standardization must be weighed against the value of locally designed authentic assessments. Along with more efforts to provide nationally normed assessment processes, the last outcome area of this domain, incorporating sociocultural factors in scientific inquiry (2.5), is rarely mentioned in the literature and needs significant attention going forward.

Several new directions in the assessment of scientific reasoning are promising. One current development in the use of standardized testing is the inclusion of the MCAT-style questions in some psychology textbook test banks (Stevens et al., 2016). These questions generally take the form of passage sets, in which students read and answer questions about a research study, including references to graphs and tables. These "MCAT scenarios" align well with the APA learning outcomes in scientific reasoning and critical thinking and can be used in the classroom setting.

Another likely development is the greater use of institutional assessment technology, for example, the move to attach assessment tools to college-wide online platforms. At a community college in Washington State where one of the authors (H.T.) teaches, the entire campus is incorporating online general education assessment rubrics for most classes (https://catalog.bellevuecollege.edu/content.php?catoid=6&navoid=186). Two of the adopted general education categories with associated scoring rubrics for relevant assignments are "Research/Information Literacy" and "Scientific Inquiry (Nature of Science)," which lend themselves well to many psychology assignments. Instructors can quickly do a four-level proficiency assessment apart from the usual grading rubric. Although students do not necessarily see the assessment scores, institutions can harvest the information for campus-wide accreditation reporting.

Finally, although the above examples of summative and formative assessments represent a strong foundation for individual and program assessment of Goal 2 outcomes, gaps remain. In some cases, appropriate assessments may have been developed but are not widely available. As mentioned above, the Summit on National Assessment in Psychology established a repository of best practices and tools for assessing student learning which is now available at Project Assessment (http://pass.apa.org). The success of the repository will depend on further growth through ongoing additions to the collection, including those assessments that have been created but have yet to be widely shared. For example, Outcome 2 measures of information literacy at Project Assessment include an assessment which addresses locating appropriate sources but is limited in assessments addressing other critical elements of information literacy, such as effective search strategies, evaluating the relevancy, currency, authority, and accuracy of information found and determining when sufficient information has been located to answer the question under investigation. Yet many faculty have developed strong literature review assignments in their courses which encompass all of the aforementioned components of information literacy. Thus, we would encourage faculty to contribute to the field by adding such assignments to public collections of assessments such as Project Assessment.

In other cases, better and more varied assessments need to be created. In particular, constructing assessments as both tools for measurement of student learning and tools to foster further student learning can lead to the integration of curriculum and assessment through the development of formative assessments that systematically build toward complex student learning captured in meaningful summative assessments. Mueller and Coon (2013) developed a series of brief assignments to build specific scientific thinking skills that students then summatively demonstrated through papers and essay exam questions. We can further ask, how can students carefully and consistently develop the understanding and skills to demonstrate psychology information literacy, or the capacity to interpret, design, and conduct basic psychological research, or incorporate sociocultural factors in scientific inquiry. Furthermore, we can ask how these efforts can become part of a nationally consistent and coordinated assessment process tied to the Guidelines. As faculty of psychology wrestle with these questions and generate creative answers to them, the discipline of psychology will be even better positioned to foster and document student growth in the foundational outcomes of scientific inquiry and critical thinking.

Acknowledgments

The authors are grateful for the American Psychological Association's (APA's) Education Directorate and Committee on Associate and Baccalaureate Education (CABE) sponsorship of SNAP.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Additional resources were provided by grants from the National Science Foundation (Division of Undergraduate Education, award number 1622982), the University of Wisconsin–Green Bay, APA's Board of Educational Affairs, Society for the Teaching of Psychology, Psi Chi, and Psi Beta.

Note

1. Deepfakes are "highly realistic and difficult-to-detect digital manipulations of audio or video" which almost anyone can now create to suggest that people have said or done things that never really happened (Citron, 2018, para. 3).

References

Adam, A., & Manson, T. M. (2014). Using a pseudoscience activity to teach critical thinking. *Teaching of Psychology*, *41*(2), 130–134. https://doi-org.ezproxy.library.ewu.edu/10.1177/0098628314530343

American Psychological Association. (2011). *Principles for quality undergraduate education in psychology*. https://www.apa.org/education/undergrad/principles-undergrad.pdf

American Psychological Association. (2013). *APA guidelines for the undergraduate psychology major: Version 2.0*. http://www.apa.org/ed/precollege/about/psymajor-guidelines.pdf

American Psychological Association. (2016). Guidelines for the undergraduate psychology major: Version 2.0. *American Psychologist*, *71*(2), 102–111. http://dx.doi.org/10.1037/a0037562

American Psychological Association. (2017). *Multicultural guidelines: An ecological approach to context, identity, and intersectionality*. http://www.apaorg/about/policy/multiculturalguidelines.pdf

Bailey, K. G. D., Rembold, L., & Abreu, C. M. (2019). Critical thinking dispositions and skills in the undergraduate research methods classroom. *Scholarship of Teaching and Learning in Psychology*. https://doi.org/10.1037/stl0000158

Bandura, A., Ross, D., & Ross, S. A. (1961). Transmission of aggression through imitation of aggressive models. *Journal of Abnormal and Social Psychology*, *63*(3), 575–582.

Bartels, J. M., & Griggs, R. A. (2019). Using new revelations about the Stanford prison experiment to address APA undergraduate psychology major learning outcomes. *Scholarship of Teaching and Learning in Psychology*. https://doi.org/10.1037/stl0000163

Baumeister, R. F., DeWall, C. N., Ciarocco, N. J., & Twenge, J. M. (2005). Social exclusion impairs self-regulation. *Journal of Personality and Social Psychology*, *88*(4), 589–604. https://doi.org/10.1037/0022-3514.88.4.589

Bensley, D. A., & Lilienfeld, S. O. (2015). What is a psychological misconception? Moving toward an empirical answer. *Teaching of Psychology*, *42*(4), 282–292. http://dx.doi.org/10.1177/0098628315603059

Blessing, S. B., & Blessing, J. S. (2010). Psych busters: A means of fostering critical thinking in the introductory course. *Teaching of Psychology*, *37*(3), 178–182. https://doi-org.ezproxy.library.ewu.edu/10.1080/00986283.2010.488540

Citron, D. (2018). *Deepfakes and the new disinformation war*. The Center for Internet and Society. http://cyberlaw.stanford.edu/publications/deepfakes-and-new-disinformation-war

Gurung, R. A. R., Hackathorn, J., Enns, C., Frantz, S., Cacioppo, J. T., Loop, T., & Freeman, J. E. (2016). Strengthening introductory psychology: A new model for teaching the introductory course. *American Psychologist*, *71*(2), 112–124. https://doi-org.ezproxy.library.ewu.edu/10.1037/a0040012

Haladyna, T. M. (1999). *Developing and validating multiple-choice test items*. Lawrence Erlbaum Associates.

Lawson, T. J., & Brown, M. (2018). Using pseudoscience to improve introductory psychology students' information literacy. *Teaching of Psychology*, *45*(3), 220–225. http://dx.doi.org/10.1177/0098628318779259

Lilienfeld, S. O., Lynn, S. B., Rusci, J., & Beyerstein, B. L. (2010). *Fifty great myths of popular psychology: Shattering widespread misconceptions about human behavior*. Wiley-Blackwell.

Loftus, E., & Palmer, J. C. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior*, *13*(5), 585–589. https://doi.org/10.1016/S0022-5371(74)80011-3

Marek, P., Christopher, A. N., & Walker, B. J. (2004). Learning by doing: Research methods with a theme. *Teaching of Psychology*, *31*(2), 128–131. https://search-ebscohostcom.ezproxy.library.ewu.edu/login.aspx?direct=true&db=psyh&AN=200413956-016&site=ehost-live&scope=site

Muehlenkamp, J. J., Weiss, N., & Hansen, M. (2015). Problem-based learning for introductory psychology: Preliminary supporting evidence. *Scholarship of Teaching and Learning in Psychology*, *1*(2), 125–136. https://doi-org.ezproxy.library.ewu.edu/10.1037/stl0000027.supp (Supplemental)

Mueller, J. F. (2019). *What is authentic assessment? Authentic assessment toolbox*. http://jfmueller.faculty.noctrl.edu/toolbox/whatisit.htm

Mueller, J. F., & Coon, H. M. (2013). Undergraduates' ability to recognize correlational and causal language before and after explicit instruction. *Teaching of Psychology*, *40*(4), 288–293. http://dx.doi.org/10.1177/0098628313501038

Norcross, J. C., Hailstorks, R., Aiken, L. S., Pfund, R. A., Stamm, K. E., & Christidis, P. (2016). Undergraduate study in psychology: Curriculum and assessment. *American Psychologist*, *71*(2), 89–101. https://doi-org.ezproxy.library.ewu.edu/10.1037/a0040095

Penningroth, S. L., Despain, L. H., & Gray, M. J. (2007). A course designed to improve psychological critical thinking. *Teaching of Psychology*, *34*(3), 153–157. http://dx.doi.org/10.1080/00986280701498509

Popham, W. J. (1999). Why standardized tests don't measure educational quality. *Educational Leadership*, *56*(6), 8–15.

Sibulkin, A. E., & Butler, J. S. (2019). Learning to give reverse causality explanations for correlations: Still hard after all these tries. *Teaching of Psychology*, *46*(3), 223–229. http://dx.doi.org/10.1177/0098628319853936

Stevens, C., & Witkow, M. R. (2014). Training scientific thinking skills: Evidence from an MCAT^2-aligned classroom module. *Teaching of Psychology*, *41*(2), 115–121. https://doi.org.ezproxy.library.ewu.edu/10.1177/0098628314530341

Stevens, C., Witkow, M. R., & Smelt, B. (2016). Strengthening scientific reasoning skills in introductory psychology: Evidence from community college and liberal arts classrooms. *Scholarship of Teaching and Learning in Psychology*, *2*(4), 245–260. https://doi.org/ezproxy.library.ewu.edu/10.1037/stl0000070

Walvoord, B. E., & Pool, K. J. (1998). Enhancing pedagogical productivity. In J. E. Groccia & J. E. Miller (Eds.), *Enhancing productivity: Administrative, instructional, and technological strategies* (pp. 35–48). Jossey-Bass.

Wentworth, D. K., & Whitmarsh, L. (2017). Thinking like a psychologist introductory psychology writing assignments: Encouraging critical thinking and resisting plagiarism. *Teaching of Psychology*, *44*(4), 335–341. http://dx.doi.org/10.1177/0098628317727909