Scholarship of Teaching and Learning in Psychology

Exemplars of the Gold Standard in SoTL for Psychology

Georjeanna Wilson-Doenges, Jordan D. Troisi, and Robert A. Bartsch

TEACHER-READY RESEARCH REVIEW

Exemplars of the Gold Standard in SoTL for Psychology

Georjeanna Wilson-Doenges
University of Wisconsin—Green Bay

Jordan D. Troisi
Sewanee: The University of the South

Robert A. Bartsch
University of Houston—Clear Lake

In a 2013 article, Wilson-Doenges and Gurung made a case for the use of benchmarks for the scholarship of teaching and learning (SoTL), setting aspirational goals for researchers (i.e., a gold standard). In order for SoTL to embody the rigors associated with other scientific research, and for the research to earn the same respect as other psychology discipline-specific research, following these standards of research methodologies is key. These standards include the following: theory-based hypotheses, longitudinal designs, true experimental designs, large sample sizes with established power, samples from more than 1 institution, advanced and multivariate data analyses, high standards of ethics, and mixed-method approaches. In this article we provide exemplars of published SoTL research in psychology that achieve this gold standard of research quality. These exemplars allow future researchers to make aspirational goals more achievable through the real-life demonstration of possibilities.

Keywords: SoTL, research methods

In a 2013 article, Wilson-Doenges and Gurung made a case for the use of benchmarks for the scholarship of teaching and learning (SoTL), setting aspirational goals for researchers (Wilson-Doenges & Gurung, 2013). SoTL has a solid history in the psychology discipline (e.g., Griggs & Collisson, 2013; Gurung, Ansburg, Alexander, Lawrence, & Johnson, 2008; Gurung & Schwartz, 2013), although the methodology used to accomplish the research has varied. In order for SoTL to embody the rigors associated with other scientific research and for SoTL research to earn the same respect as other psychology discipline-specific research, follow-

ing the "gold standard" of research methodologies is key. Other disciplines have noted individual standards that should be pursued. Most common is the gold standard of random assignment with researchers arguing both for (e.g., Dewar, 2012; Mosteller & Boruch, 2002) and against its necessity (e.g., Sullivan, 2011). Although other disciplines have discussed individual standards, we did not find any other coherent set of gold standards for SoTL researchers.

SoTL is more than just assessing student preferences or learning. Potter and Kustra (2011, p. 2) defined SoTL as "the systematic study of teaching and learning, using established or validated criteria of scholarship, to understand how teaching (beliefs, behaviors, attitudes, and values) can maximize learning, and/or develop a more accurate understanding of learning, resulting in products that are publicly shared for critique and use by an appropriate community." SoTL necessitates that instructors make intentional changes in the teaching/ learning experience and then assess how effective that change was in improving the learning of their students. This type of language

Georjeanna Wilson-Doenges, Department of Psychology and Human Development, University of Wisconsin—Green Bay; Jordan D. Troisi, Department of Psychology, Sewanee: The University of the South; Robert A. Bartsch, Department of Psychology, University of Houston—Clear Lake.

Correspondence concerning this article should be addressed to Georjeanna Wilson-Doenges, 2420 Nicolet Drive, MAC C310, University of Wisconsin—Green Bay, Green Bay, WI 54311. E-mail: wilsong@uwgb.edu

is also used to describe other more traditional empirical psychological research. The gold standard of SoTL should be "theory based, have established power, use reliable and valid measures, use robust methodologies, and utilize advanced and multivariate techniques to analyze properly screened data" (Wilson-Doenges & Gurung, 2013, p. 68). The goal of using gold standard methodologies is to create a "greater likelihood that trustworthy information will be created and disseminated" (Potter & Kustra, 2011, p. 2). As evidence accumulates using these effective and advanced techniques, teachers need no longer to rely on hunches about student learning.

In this article we provide exemplars of published SoTL research in psychology that achieve elements of this gold standard of research quality within the context of teaching and learning (see Table 1 for a summary). We identified these exemplars from a review of recent publications in SoTL journals in psychology (e.g., *Teaching of Psychology, Scholarship of Teaching and Learning in Psychology*), identifying studies that exemplified aspects of the gold standard (e.g., used samples taken from more than one institution with some diversity). We hope these exemplars will make aspirational goals more achievable through demonstrating

how other researchers have met these goals. Although it is unlikely that any study can meet all the standards at once, quality research should strive to meet several of these benchmarks to enhance the contribution to the field. Some of the benchmarks of the gold standard may be more important than others at a particular point in time or for a particular study. For example, for topics that imply long-term changes (e.g., reduction in prejudice), longitudinal students may be exceptionally valuable. But for studies that require causality to be affirmed (e.g., does a specific activity lead to increased engagement?), experimental studies may be important. Both of these are gold standards to be achieved, and the context and topic inform what is more important to strive for. The importance may also depend on the state of the literature. If there are no studies with good power in the literature, a high-power study may be needed. Therefore, identifying particular benchmarks to strive for may be based on the study being pursued or the state of the field at the time.

Gold Standard Benchmark 1: Theory Based

A strong literature backing can provide a solid foundation for the design of a SoTL re-

Table 1
Summary of Exemplars of Gold Standards in SoTL Research

Gold standard	Main characteristic	Exemplar
Theory based	Bases work on developed theories and past research.	Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013; Troisi (2015)
Longitudinal design	Measures participants over time with good response rates and tracks change scores.	Buch & Spaulding (2011); Kernahan & Davis (2010)
True experimental design	Random assignment with double-blind procedures.	Balch (2012); Legg & Wilson (2009); Poonati & Amadio (2010)
Large sample sizes	Has high statistical power.	Thompson & Fisher-Thompson (2013); Troisi (2014); Warren (2006)
Diverse samples	Samples taken from more than one institution with some diversity.	Gurung, Daniel, & Gurung (2012); Homa et al. (2013); Howell, Collisson, & King (2014); Troisi, Leder, Stiegler-Balfour, Fleck, & Good (2015)
Advanced statistical techniques	Proper data screening; appropriate advanced (often multivariate) statistical analyses.	Preckel et al. (2013); Renken, McMahan, & Nitkova (2015); Rogers (2015)
High standard of ethics	Students not coerced into participating, confidentially maintained; just access to benefits to all participants.	Franz & Spitzer (2006); Funk & Dickson (2011); Ocker & Yaverbaum (1999)
Mixed-method approach	Using both qualitative and quantitative data analyzed appropriately.	Bridges, Harnish, & Sillman (2012); Knott, Mak, & Neill (2013); Martin (2015)

Note. The list of gold standards is from Wilson-Doenges and Gurung (2013). SoTL = scholarship of teaching and learning.

search project that meets this gold standard benchmark. Utilizing strengths and improving upon weaknesses of previously published studies, researchers can ground their independent and dependent variables in a rich theoretical and empirical history. Using previously established reliable and valid methods and measures can improve the quality of the research design and, ultimately, one's results and conclusions.

In one such example, Troisi (2015) utilized a pretest–posttest quasi-experimental design to demonstrate that the use of student management teams in the classroom produced improvements in student autonomy. Autonomy has a long-standing history as one of three fundamental needs for human motivation from self-determination theory (Deci & Ryan, 1987, 2000; Ryan & Deci, 2000a, 2000b, 2006, 2009). Previous research has also established that autonomy facilitates positive student outcomes, including more effective and longer-lasting learning, greater interest, less pressure and tension, more creativity, more cognitive flexibility, better learning, better self-control, a more positive emotional tone, and higher self-esteem (e.g., Kusurkar, Ten Cate, Vos, Westers, & Croiset, 2013; Niemiec & Ryan, 2009; Patrick, Skinner, & Connell, 1993; Ryan, Connell, & Plant, 1990; Vansteenkiste, Simons, Lens, Sheldon, & Deci, 2004; Wong, 2000). Thus, improving students' feelings of autonomy in the classroom is a desirable goal.

Drawing from this long-standing psychological theory, Troisi (2015) utilized an existing self-determination theory questionnaire for autonomy from previous research (e.g., Black & Deci, 2000; Williams & Deci, 1996; Williams, Saizow, Ross, & Deci, 1997) to empirically examine a potential autonomy bolstering pedagogical technique: the use of student management team (cf. Handelsman, 2012; Nuhfer, 1997, 2008; Schwartz, 1996). Briefly, a student management team is a group of students who work with a professor on an ongoing basis to help alter the content and delivery of a course. Presumably, this technique should allow students in a class with an embedded student management team to feel more autonomous and in control of outcomes in their course. In a quasi-experimental design, Troisi (2015) examined whether including a student management team in a course would increase students' levels of autonomy more than in a control comparison

course in which no student management team was present. The results confirmed this expectation. This study provides an example of an important psychological variable (i.e., autonomy), which is drawn from a long-standing psychological theory (i.e., self-determination theory). The study also adds a new link between pedagogical practices and feelings of student autonomy.

In another example of theory-based research, there are many studies of test-enhanced learning that are informed by well-replicated empirical work and theory (Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013). This body of research shows that retrieval through practice testing consolidates learning, illustrating that testing, not studying, increases long-term recall. Dunlosky et al. (2013) provided a summary of the evidence-based strategy of using practice testing to improve student learning. Their summary illustrates two theoretical explanations for the testing effect: direct effects that spark elaborative retrieval processes, and mediated effects for which practice testing increases encoding during later studying. This collection of research, rooted in theory, supports the notion that testing enhances memory and learning. An exemplar, informed by the vast literature on the testing effect, is McDaniel, Wildman, and Anderson's (2012) study of online quizzes used to enhance test performance in college students. They found that repeated quizzing improved exam scores under several circumstances (McDaniel et al., 2012).

Gold Standard Benchmark 2: Longitudinal Design Tracking Students Over Time With Good Response Rates and Analyses That Track Change Scores

The ability to measure the dependent variable multiple times throughout the semester and beyond can uncover changes in learning as the semester progresses and promotes understanding of when and how changes occur. This gold standard methodology necessitates that response rates remain high over the course of the data collection and that analyses are appropriate to understand the dynamic nature of the data. Having the ability to see whether a SoTL intervention has an immediate or delayed impact on learning and whether the effects are short or long term denotes quality information that

would inform teaching and learning, and ultimately, the goals of higher education.

An exemplar of a longitudinal study that meets this benchmark is Kernahan and Davis's (2010) tracking of the impact of taking a diversity class on awareness and attitudes over the semester and 1-year beyond. Students enrolled in a diversity class and a control group class (in statistics) were assessed on several measures of racial attitudes and awareness 1 week prior to the start of their respective classes, at the end of the semester, and 1 year later. Findings showed that students in the diversity class had changes in racial attitudes and awareness by the end of the semester, but some of those changes remained the same or diminished in the year following the class. However, one year after taking the diversity class, those students reported increased comfort regarding racial issues and interactions with other races. Although the response rate over the three data collections did decrease, a majority of the students were retained for all three. This kind of methodology allows instructors to understand the impact of a class immediately and also in the longer term, helping instructors know the longevity of the impact made on students.

Another study that exemplifies the gold standard of longitudinal research is the 2011 study of Psychology Learning Communities (PLC) by Buch and Spaulding. They tracked students in six cohorts longitudinally over their college careers, starting in their first semester with matched comparison groups for each cohort. Buch and Spaulding (2011) found that students who participated in a PLC had significantly better first-semester grade-point averages (GPAs), retention, and progression toward their degree, and were significantly more involved in the psychology major. Furthermore, their longitudinal method with comparison groups allowed these researchers to understand the long-term impact of SoTL projects on student outcomes such as retention, GPA, and engagement.

Gold Standard Benchmark 3: True Experimental Design That Includes Random Assignment and Double-Blind Procedures

Because SoTL happens in large part within the context of a class, this benchmark is often

difficult to accomplish. Randomly assigning students to certain classes or having multiple conditions within a single class can seem like an unachievable goal. However, true double-blind experiments are the highest form of controlled empirical research in the field of psychology, and as such, set a high goal for SoTL research. Experimental controls allow researchers to rely on the comparisons made with the control group to truly know that the treatment was a factor in increased student learning.

In one such example, Balch (2012) utilized an experimental methodology in an introductory psychology course during which all students took a pretest and then were randomly assigned to two groups. The two groups were asked to come to class at different times the following class period, during which both groups were presented with the same material but with different instruction methods: one by free-recall demonstration and the other by lecture. Although pretest scores did not differ between the two instructional methods, students in the free-recall demonstration condition scored higher 5 days later on a posttest, resulting in an Instructional Method $\times$ Pretest–Posttest interaction. This study is a helpful example of how random assignment can work within the context of a single class, namely, by inviting students to come to class at different times and receive different levels of the independent variable.

In another example of a true, double-blind experimental design, Legg and Wilson (2009) studied how sending a welcome e-mail to students in a class a week prior to the semester starting may affect rapport, student motivation, and retention. In this study, students were randomly assigned to either receive or not receive a welcome e-mail, and their attitudes were tracked over the semester. The instructor was unaware of the condition to which the students had been assigned, because the graduate assistant sent the e-mail to the randomly selected half of the class using the instructor's e-mail address. Students who received the e-mail reported significantly better motivation and attitudes toward the instructor and the course on the first day of class. Retention was also better for students who received the e-mail. This study is a particularly good exemplar for the ways in which researchers can study students in a single class with random assignment and double-blind conditions.

As a third example, Poonati and Amadio (2010) brought a classroom situation into the laboratory to determine whether watching videos from popular TV shows would help students learn concepts related to operant conditioning. Having the study in the laboratory allowed researchers to create a true experiment by easily randomly assigning participants to either see the videos containing exemplars or have the same exemplars verbally described.

Gold Standard Benchmark 4: Large Sample Sizes and Established Power of Statistical Test

This is one of the most difficult benchmarks to achieve—especially for those teaching at small institutions or those who have small class sizes. Larger sample sizes help meet assumptions of rigorous statistical tests and increase the power of those tests. However, obtaining large sample sizes is not always possible or practical. Tomcho and Foels (2009) discussed the challenges of achieving appropriate power within the SoTL context and gave some recommendations for ways to achieve better power. Although there are some wonderful exemplars of large sample size studies in SoTL (e.g., Warren, 2006), we use some exemplars of ways to achieve the most power within the context of somewhat smaller sample sizes.

Bartsch, Engelhardt Bittner, and Moreno (2008) proposed a methodology that would improve power even when sample sizes are small. Their design included two versions of the assessments that allowed a pretest–posttest experimental methodology while still retaining the whole class size instead of dividing into two groups. Thompson and Fisher-Thompson (2013) later used this methodology. They wanted to determine whether an activity that included watching a brief video about a research project and then analyzing and interpreting example data from that project helped students better understand statistics. If they assigned their 30 students into groups that received the activity and those that did not, their analysis would have had half the participants and hence, lower power. Instead, they created two versions of the assessment. Half the students received Version A as a pretest and Version B as a posttest, and the other half received Version B as a pretest and Version A as a posttest. This manipulation was powerful because it used a repeated-measures design that used all participants in both conditions to maintain a larger sample size. Within-group designs are beneficial in that they can use all participants in all conditions and increase power in that way.

In another example of increasing power within the context of smaller sample sizes, Troisi (2014) used a bootstrapping analysis in his pretest–posttest quasi-experimental examination of the effects of student membership on a student management team in predicting course engagement and performance. The bootstrapping technique estimates the relationships between conceptual variables by treating sample data as a population, then resampling those data numerous times in order to provide estimates of direct and indirect effects, and ultimately, estimate the population indirect effect, standard error, and lower and upper confidence intervals of the indirect effect (Preacher & Hayes, 2008; Shrout & Bolger, 2002). The bootstrapping technique is particularly useful for underpowered designs, such as designs with small samples (MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002; Preacher & Hayes, 2008; Shrout & Bolger, 2002), making it a potentially effective tool in SoTL research studies involving small groups of participants. Results of Troisi's (2014) study revealed that management team participation produced an increase in course performance and course engagement, as well the finding that course engagement mediated the increase in course performance. This effect was evidenced despite the fact that one cell in the 2×2 quasi-experimental design had only 6 participants (those who were members of a student management team in two courses). Bootstrapping analyses allow for rigorous, multivariate tests of hypotheses with small samples. Their tools can also be freely downloaded and installed within traditional statistical software programs (e.g., Hayes, n.d.).

Gold Standard Benchmark 5: Samples Taken From More Than One Institution With Some Diversity

Although instructors can learn much from studying their own classes at their own institutions, external validity would be vastly improved by studying the same phenomena in large and diverse samples from multiple insti-

tutions. Broadening samples to include students from national or international populations can help researchers understand how learning can be impacted in multiple contexts.

In one such study, Gurung, Daniel, and Landrum (2012) examined learning outcomes among 454 students from at least three different schools (two large public institutions, one small liberal arts school). The conclusions that they were able to draw show that students' metacognitive skill, use of pedagogical aids to study, and total study time had more impact on quiz scores than did teacher behavior or textbook quality ratings. The value of these findings are made even more relevant because they come from a large and diverse sample of students using different textbooks and learning from different instructors rather than using just one text or instructor. This diverse sample helps rule out important confounds such as idiosyncratic tendencies of particular instructors, particular courses, or particular institutions.

Another exemplar of a multi-institutional study is Homa et al.'s (2013) examination of student learning objectives (SLOs) and course content in introductory psychology courses using 158 syllabi from 95 institutions. The results from their content analysis showed that more than 50% of the syllabi listed objectives that mapped onto the science and application of psychology SLOs, and that cognitive and physiological topics were covered in more depth than were history, research methods, and developmental psychology. Using a national sample of syllabi helps instructors understand the broad scope of coverage in introductory psychology courses that a smaller, less-representative sample could not achieve. Technology, such as the listservs and Websites used in this study, allows researchers to access institutions nationwide fairly easily, allowing sample diversity, especially when looking at course syllabi. We encourage research that uses Web-based tools to collaborate with other researchers or collect diverse samples.

Troisi, Leder, Stiegler-Balfour, Fleck, and Good (2015) used online surveying to examine mentorship outcomes of 122 Early Career Psychologist faculty members. This sample was recruited from major psychology professor groups (e.g., the PSYCHTEACHER listserv, the Early Career Psychologists Committee Facebook page), yielding a sample with diverse professional experiences. For example, the sample included professors from a wide range of academic settings (e.g., 4-year liberal arts colleges, research-intensive universities, professional schools) who specialized in a variety of subfields of psychology (e.g., social, cognitive, developmental). Such diverse samples are increasingly available with the proliferation of Web-based surveying tools (e.g., Qualtrics, SurveyMonkey) as well as with Web-based tools for recruiting participants at low cost (e.g., e-mail lists, social media outlets, Amazon's Mechanical Turk).

Lastly, one of the best examples of diverse samples is Howell, Collisson, and King's (2014) worldwide sample of 209 psychologists and 132 physicists in a study of perceptions of the nature and core content of their respective fields. Accessing academic institutions from the United States, Canada, Europe, Australia, and New Zealand, they found that psychologists thought that psychology was less grounded in theory and empirical evidence with little consensus about the field's core content in comparison with physicists. A global sample is truly reaching this gold standard benchmark, allowing instructors to assess opinions from a diverse set of institutions around the world.

Gold Standard Benchmark 6: Proper Data Screening and Statistical Analyses That Are Advanced and Multivariate

Proper data screening can help ensure that data meet the assumptions of the statistical tests used in hypothesis testing and inferential statistics. In addition, data screening helps researchers understand the shortcomings of their data, such as outliers or missing data. Violation of assumptions of the statistics used can limit the reliability and the power of results. Although very important, data screening is usually something that is completed before hypotheses are tested and is often not reported in published research. Data screening is crucial to the research process, but is largely assumed rather than reported in many published studies. Another part of data screening, in the larger sense, is testing the psychometric properties of any scales used in analyses. Ensuring that any scales are reliable and valid is an important consideration before testing hypotheses.

One example of such scale validation is Renken, McMahan, and Nitkova's (2015) initial validation of the Psychology-Specific Epistemological Belief Scale (Psych-SEBS). In this two-part study, researchers first drafted and then refined an item pool using exploratory factor analysis, followed by confirmatory factor analysis and an assessment of internal consistency, test–retest reliability, and convergent validity of the Psych-SEBS. In the second study, researchers assessed the criterion validity by comparing the Psych-SEBS with an established criterion and also tested the incremental validity of the Psych-SEBS above and beyond the variance explained by that criterion. By all measures, the Psych-SEBS 13-item scale is a reliable and valid measure of psychology-specific epistemological beliefs.

In addition to proper data screening, utilizing the best and most applicable multivariate statistics available can improve understanding of the complex reasons why some students learn well while others do not. Building models that test multiple variables simultaneously can help explain the complexities in the teaching and learning process.

In an extensive study of time-of-day preference (chronotype) and academic achievement, Preckel and colleagues (2013) used multiple predictors in hierarchical analysis that statistically controlled for variables suggested to be predictive of academic performance (e.g., conscientiousness, cognitive ability, achievement motivation, gender) to illuminate the incremental validity of chronotype on GPA. Before testing hypotheses Preckel et al. screened their scales and found that all scales used for prediction were reliable and bivariately correlated with the dependent variables before they were used in the regression model. Also, as a byproduct of this prescreening of scales, a suppressor effect was found between daytime sleepiness and chronotype, and therefore daytime sleepiness was excluded from the analyses. In this way, the prescreening of the scales led to a better choice in hypothesis testing by excluding daytime sleepiness. Good use of statistical control and use of multiple predictors including self- and other-report data elevate this study's methodology as an exemplar of multivariate statistics and establishment of the psychometric properties of the scales used in analyses.

Rogers' (2015) study of the Learning Alliance Inventory is another example of scale development, reliability and validity testing, and then well-executed multivariate statistics. One of the strengths of this study is the extensive psychometric scale testing, including internal and test–retest reliability, bivariate correlation, and predictive validity. Another strength is the well-designed, executed, and reported results from the path analysis establishing support for immediacy and rapport indirectly impacting learning through work alliance. Path analysis allows for testing multiple relationships simultaneously, and the findings of this research illuminate how several variables simultaneously impact learning, both directly and indirectly. In addition to quality of the analyses themselves, the tables are clearly written and provide a model for quality reporting of scale development and path analysis.

Gold Standard Benchmark 7: High Standard of Ethics

There has been some controversy over the ethics of studying students in the classroom. Making sure that SoTL research lives up to the same high ethical standards of other psychological research is paramount (e.g., Martin, 2013). Of particular concern is the issue of coercion. When students are asked to participate in a study during a class for which they are earning a grade, they could feel pressure to complete the study. There are ways for researchers to reduce the risk of coercing students to participate in research. For example, Funk and Dickson (2011) provided extra credit for correct answers to additional questions on exams. Their institutional review board approved giving extra credit because the incentive was "voluntary and nonpunitive" (Funk & Dickson, 2011, p. 272). In this way, students received a benefit for extra work, and that work did not impact their grade.

Another way to address the issue of coercion is to use assessments that are part of the regular course assignments, and then assess them blindly for the SoTL part of the project separately from the assignment grade for course credit. In order for this to work though, students must believe that their identity is completely anonymous and that the instructor is unaware of the condition when grading. An example of this type of study is Franz and Spitzer's (2006)

study of different approaches to teaching American Psychological Association (APA) style. In their study they assessed the APA style papers that students wrote as part of their regular course assignments after being exposed to a checklist approach, a template approach, or a combination of the two. They maintained the anonymity of their students by having the students use pseudonyms on all materials in the class. The instructor was also blind to the condition of type of method the student used in preparing the APA style paper because the lab assistants coded the papers and then randomized their order before the instructor evaluated them. Using pseudonyms and having someone other than the instructor code assignments keeps the researcher unaware of the condition and protects the student from possible repercussions on their grades. This allows grades and SoTL assessments to remain separate, reducing the pressure that students may feel to participate.

Providing equivalent and just access to benefits for all participants is also imperative. This may seem impossible in a short semester-long time span, but there are ways to accomplish justice in SoTL work. Smith (2008) suggested an instructor split the class and have the experimental group meet during the first part of class and the control group meet during the second part. After a formative assessment, the instructor could determine whether the intervention given to the experimental group was beneficial and, if so, provide the intervention to the control group before any grading of the material. This is a way to quickly assess the benefits and share those benefits with the control group before the end of the semester.

Ocker and Yaverbaum (1999) had another approach to justice in their examination of whether face-to-face or asynchronous online communication was better for group case studies. They evaluated two cases in two different sections. The first section did Case Study A communicating face to face and Case Study B online. The second section did Case Study A online and Case Study B face to face. This design, which could have been done within a single class, equalized any modality benefits between the two classes. In other words, no group would have received an overall benefit or harm in comparison with the other groups, thus maintaining justice.

Gold Standard Benchmark 8: Mixed-Method Approach—Using Both Qualitative and Quantitative Data Analyzed Appropriately

Using both high-quality qualitative and quantitative data analysis can help add to a deeper understanding of teaching and learning. Qualitative data that have been properly analyzed (perhaps with software like NVivo, 2012, or others) can incorporate the voices of students better than the more standardized assessment using quantitative measures. It is important to be mindful to use qualitative data in a way that maintains anonymity and confidentiality, even when reported in the participant's own words. Hand in hand, these two methodologies can provide researchers with rich data to further understand learners' experiences.

One exemplary study using a mixed-method approach to data analysis is Knott, Mak, and Neill's (2013) study of intercultural competency training through skill development. Participants completed an online survey with both quantitative items and one open-text question. The quantitative results revealed that participants felt that the training in introductory psychology had increased their cultural competence. However, it is the qualitative analysis that really reveals the richness of the changes in competency as seen through their open-text responses to applications of intercultural competency. The researchers followed the Attride-Stirling (2001) method of systematically analyzing their qualitative data into a hierarchy of themes, which included global, organizing, and basic themes. This method revealed one global theme that encapsulated two organizing themes and six basic themes, showing in-depth ways in which participants had applied what they learned in their training to demonstrate cultural competence.

As another strong example of the mixed-method approach, Martin (2015) performed an exploratory analysis of the qualities of exceptional undergraduate students in psychology. He surveyed professors, using open-ended questions to identify the top three qualities of their top students (e.g., intelligence, strong work ethic, intellectual curiosity). These responses were then independently coded into categories by one professor and four undergraduate stu-

dents, using discussion to reconcile differences. The analysis also included Likert-type questionnaire items to quantify their abilities (e.g., "How easy was that undergraduate student to teach?") and their engagement in academic activities (e.g., honor societies, research assistantships, honors projects). Especially for an exploratory study for which there has been little background work, we perceive this mixed-method approach to be ideal.

In a third example, Bridges, Harnish, and Sillman (2012) investigated the effectiveness of using blog assignments in a positive psychology class. They used quantitative student ratings taken at the conclusion of the class. Ratings were significantly higher in the class with the blog exercise than in the previous class, which instead had a collaborative research project. In addition they examined students' attitudes toward the course in an open-ended survey 6 months after the course finished. They used previously published methods (i.e., Sherman & MacDonald, 2009) as a pattern for qualitatively identifying themes. These themes indicated that they enjoyed the class and the blogging assignments, they continued to use information they learned while doing the assignments, but most had not continued to blog.

Conclusion

Following the gold standard benchmarks of methodologies in SoTL will ensure that teachers have the best evidence-based information available to them to improve the quality of student learning. Using methodologies that follow the highest standards will make better teachers and will ensure that the information researchers disseminate about teaching and learning is of the highest quality. As standards in the field constantly evolve, it is important to stay current with the best practices in the field. One way to accomplish this is keeping abreast of new books and articles as they are published (e.g., Jhangiani, Troisi, Fleck, Legg, & Hussey, 2015, on tools and measures for SoTL). Sometimes achieving meaningful, theoretically driven, and methodologically rigorous research in SoTL may seem impossible. This article provides exemplars of benchmarks of the gold standard in SoTL research as set out by Wilson-Doenges and Gurung (2013) to show how these

standards are achievable for all SoTL researchers.

References

Attride-Stirling, J. (2001). Thematic networks: An analytical tool for qualitative research. *Qualitative Research, 1,* 385–405. http://dx.doi.org/10.1177/146879410100100307

Balch, W. R. (2012). A free-recall demonstration versus a lecture-only control: Learning benefits. *Teaching of Psychology, 39,* 34–37. http://dx.doi.org/10.1177/0098628311430170

Bartsch, R. A., Engelhardt Bittner, W. M., & Moreno, J. E., Jr. (2008). A design to improve internal validity of assessments of teaching demonstrations. *Teaching of Psychology, 35,* 357–359. http://dx.doi.org/10.1080/00986280802373809

Black, A. E., & Deci, E. L. (2000). The effects of instructors' autonomy support and students' autonomous motivation on learning organic chemistry: A self-determination theory perspective. *Science Education, 84,* 740–756. http://dx.doi.org/10.1002/1098-237X(200011)84:6<740::AID-SCE4>3.0.CO;2-3

Bridges, K. R., Harnish, R. J., & Sillman, D. (2012). Teaching undergraduate positive psychology: An active learning approach using student blogs. *Psychology Learning & Teaching, 11,* 228–237. http://dx.doi.org/10.2304/plat.2012.11.2.228

Buch, K., & Spaulding, S. (2011). The impact of a psychology learning community on academic success, retention, and student learning outcomes. *Teaching of Psychology, 38,* 71–77. http://dx.doi.org/10.1177/0098628311401589

Deci, E. L., & Ryan, R. M. (1987). The support of autonomy and the control of behavior. *Journal of Personality and Social Psychology, 53,* 1024–1037. http://dx.doi.org/10.1037/0022-3514.53.6.1024

Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry, 11,* 227–268. http://dx.doi.org/10.1207/S15327965PLI1104_01

Dewar, J. M. (2012). Gathering and analyzing evidence. In J. M. Dewar & C. D. Bennett (Eds.), *Doing the scholarship of teaching and learning in mathematics* (pp. 19–44). Washington, DC: Mathematical Association of America.

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, 14,* 4–58. http://dx.doi.org/10.1177/1529100612453266

Franz, T. M., & Spitzer, T. M. (2006). Different approaches to teaching the mechanics of APA style. *Journal of the Scholarship of Teaching and Learning, 6,* 13–20.

Funk, S. C., & Dickson, K. L. (2011). Multiple-choice and short-answer exam performance in a college classroom. *Teaching of Psychology, 38,* 273–277. http://dx.doi.org/10.1177/0098628311421329

Griggs, R. A., & Collisson, B. (2013). ToP's greatest hits: The most frequently cited *Teaching of Psychology* articles. *Teaching of Psychology, 40,* 76–87. http://dx.doi.org/10.1177/0098628312475026

Gurung, R. A. R., Ansburg, P. I., Alexander, P. A., Lawrence, N. K., & Johnson, D. E. (2008). The state of scholarship of teaching and learning in psychology. *Teaching of Psychology, 35,* 249–261. http://dx.doi.org/10.1080/00986280802374203

Gurung, R. A. R., Daniel, D. B., & Landrum, R. E. (2012). A multisite study of learning in introductory psychology courses. *Teaching of Psychology, 39,* 170–175. http://dx.doi.org/10.1177/0098628312450428

Gurung, R. A. R., & Schwartz, B. M. (2013). Contributions from psychology: Heuristics for interdisciplinary advancement of SoTL. In K. McKinney (Ed.), *The scholarship of teaching and learning in and across the disciplines* (pp. 34–53). Bloomington: Indiana University Press.

Handelsman, M. M. (2012). Course evaluation for fun and profit: Student management teams. In J. Holmes, S. C. Baker, & J. R. Stowell (Eds.), *Essays from e-xcellence in teaching* (pp. 8–11). Retrieved from http://teachpsych.org/ebooks/eit2011/index.php

Hayes, A. F. (n.d.). *SPSS, SAS, and Mplus macros and code.* Retrieved from http://www.afhayes.com/spss-sas-and-mplus-macros-and-code.html

Homa, N., Hackathorn, J., Brown, C. M., Garczynski, A., Solomon, E. D., Tennial, R., . . . Gurung, R. A. R. (2013). An analysis of learning objectives and content coverage in introductory psychology syllabi. *Teaching of Psychology, 40,* 169–174. http://dx.doi.org/10.1177/0098628313487456

Howell, J. L., Collisson, B., & King, K. M. (2014). Physics envy: Psychologists' perceptions of psychology and agreement about core concepts. *Teaching of Psychology, 41,* 330–334. http://dx.doi.org/10.1177/0098628314549705

Jhangiani, R. S., Troisi, J. D., Fleck, B., Legg, A., & Hussey, H. (Eds.). (2015). *A compendium of scales for use in the scholarship of teaching and learning.* Retrieved from http://teachpsych.org/ebooks/eit2011/index.php

Kernahan, C., & Davis, T. (2010). What are the long-term effects of learning about racism? *Teaching of Psychology, 37,* 41–45. http://dx.doi.org/10.1080/00986280903425748

Knott, V. E., Mak, A. S., & Neill, J. T. (2013). Teaching intercultural competencies in introductory psychology via application of the Excellence in Cultural Experiential Learning and Leadership model. *Australian Journal of Psychology, 65,* 46–53. http://dx.doi.org/10.1111/ajpy.12008

Kusurkar, R. A., Ten Cate, T. J., Vos, C. M. P., Westers, P., & Croiset, G. (2013). How motivation affects academic performance: A structural equation modelling analysis. *Advances in Health Sciences Education, 18,* 57–69. http://dx.doi.org/10.1007/s10459-012-9354-3

Legg, A. M., & Wilson, J. H. (2009). E-mail from professor enhances student motivation and attitudes. *Teaching of Psychology, 36,* 205–211. http://dx.doi.org/10.1080/00986280902960034

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods, 7,* 83–104. http://dx.doi.org/10.1037/1082-989X.7.1.83

Martin, R. C. (2013). Navigating the IRB: The ethics of SoTL. *New Directions for Teaching and Learning, 2013*(136), 59–71. http://dx.doi.org/10.1002/tl.20076

Martin, R. C. (2015). Undergraduate superstars: What makes them stand out? *Scholarship of Teaching and Learning in Psychology, 1,* 107–112. http://0-dx.doi.org.catalog.sewanee.edu/10.1037/stl0000028

McDaniel, M. A., Wildman, K. M., & Anderson, J. L. (2012). Using quizzes to enhance summative-assessment performance in a web-based class: An experimental study. *Journal of Applied Research in Memory & Cognition, 1,* 18–26. http://dx.doi.org/10.1016/j.jarmac.2011.10.001

Mosteller, F., & Boruch, R. (Eds.). (2002). *Evidence matters: Randomized trials in educational research.* Washington, DC: Brookings Institution.

Niemiec, C. P., & Ryan, R. M. (2009). Autonomy, competence, and relatedness in the classroom: Applying self-determination theory to educational practice. *Theory and Research in Education, 7,* 133–144. http://dx.doi.org/10.1177/1477878509104318

Nuhfer, E. B. (1997). *Student management teams—The heretic's path to teaching success.* In W. Campbell & K. Smith (Eds.), *New paradigms for college teaching* (pp. 103–126). Retrieved from http://hort.ifas.ufl.edu/als5934/campbell-smith1997.pdf

Nuhfer, E. (2008). *A handbook for student management teams.* Retrieved from http://profcamp.tripod.com/New_PM_Intro_SMT.pdf (Original work published 1991)

Nvivo. (2012). (Version 10 for Windows) [Computer software]. Burlington, MA: QSR International.

Ocker, R. J., & Yaverbaum, G. J. (1999). Asynchronous computer-mediated communication versus face-to-face collaboration: Results on student learning, quality, and satisfaction. *Group Decision and Negotiation, 8,* 427–440. http://dx.doi.org/10.1023/A:1008621827601

Patrick, B. C., Skinner, E. A., & Connell, J. P. (1993). What motivates children's behavior and emotion? Joint effects of perceived control and autonomy in the academic domain. *Journal of Personality and Social Psychology, 65,* 781–791. http://dx.doi.org/10.1037/0022-3514.65.4.781

Poonati, S., & Amadio, D. M. (2010). Use of popular television to enhance students' understanding of operant conditioning. *Psychology Learning & Teaching, 9,* 25–29. http://dx.doi.org/10.2304/plat.2010.9.1.25

Potter, M. K., & Kustra, E. (2011). The relationship between scholarly teaching and SoTL: Models, distinctions, and clarifications. *International Journal for the Scholarship of Teaching and Learning, 5.* Retrieved from www.georgiasouthern.edu/ijsotl

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods, 40,* 879–891. http://dx.doi.org/10.3758/BRM.40.3.879

Preckel, F., Lipnevich, A. A., Boehme, K., Brandner, L., Georgi, K., Könen, T., . . . Roberts, R. D. (2013). Morningness–eveningness and educational outcomes: The lark has an advantage over the owl at high school. *British Journal of Educational Psychology, 83,* 114–134. http://dx.doi.org/10.1111/j.2044-8279.2011.02059.x

Renken, M. D., McMahan, E. A., & Nitkova, M. (2015). Initial validation of an instrument measuring psychology-specific epistemological beliefs. *Teaching of Psychology, 42,* 126–136. http://dx.doi.org/10.1177/0098628315569927

Rogers, D. T. (2015). Further validation of the Learning Alliance Inventory: The roles of working alliance, rapport, and immediacy in student learning. *Teaching of Psychology, 42,* 19–25. http://dx.doi.org/10.1177/0098628314562673

Ryan, R. M., Connell, J. P., & Plant, R. W. (1990). Emotions in nondirected text learning. *Learning and Individual Differences, 2,* 1–17. http://dx.doi.org/10.1016/1041-6080(90)90014-8

Ryan, R. M., & Deci, E. L. (2000a). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology, 25,* 54–67. http://dx.doi.org/10.1006/ceps.1999.1020

Ryan, R. M., & Deci, E. L. (2000b). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55,* 68–78. http://dx.doi.org/10.1037/0003-066X.55.1.68

Ryan, R. M., & Deci, E. L. (2006). Self-regulation and the problem of human autonomy: Does psychology need choice, self-determination, and will? *Journal of Personality, 74,* 1557–1586. http://dx.doi.org/10.1111/j.1467-6494.2006.00420.x

Ryan, R. M., & Deci, E. L. (2009). Promoting self-determined school engagement: Motivation, learning, and well-being. In K. R. Wentzel & A. Wigfield (Eds.), *Handbook of motivation at school* (pp. 171–196). New York, NY: Routledge.

Schwartz, R. A. (1996, January). Improving course quality with student management teams. *ASEE Prism, 1,* 9–23. Retrieved from http://www.asee.org/papers-and-publications/publications/prism/article-index/1995-1996

Sherman, A., & MacDonald, L. (2009). Service learning experiences in university science degree courses. *Innovative Higher Education, 34,* 235–244. http://dx.doi.org/10.1007/s10755-009-9110-7

Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods, 7,* 422–445. http://dx.doi.org/10.1037/1082-989X.7.4.422

Smith, R. A. (2008). Moving toward the scholarship of teaching and learning: The classroom can be a lab, too! *Teaching of Psychology, 35,* 262–266. http://dx.doi.org/10.1080/00986280802418711

Sullivan, G. M. (2011). Getting off the "gold standard": Randomized controlled trials and education research. *Journal of Graduate Medical Education, 3,* 285–289. http://dx.doi.org/10.4300/JGME-D-11-00147.1

Thompson, W. B., & Fisher-Thompson, D. (2013). Analyzing data from studies depicted on video: An activity for statistics and research courses. *Teaching of Psychology, 40,* 139–142. http://dx.doi.org/10.1177/0098628312475035

Tomcho, T. J., & Foels, R. (2009). The power of teaching activities: Statistical and methodological recommendations. *Teaching of Psychology, 36,* 96–101. http://dx.doi.org/10.1080/00986280902739743

Troisi, J. D. (2014). Making the grade and staying engaged: The influence of student management teams on student classroom outcomes. *Teaching of Psychology, 41,* 99–103. http://dx.doi.org/10.1177/0098628314530337

Troisi, J. D. (2015). Student management teams increase college students' feelings of autonomy in the classroom. *College Teaching, 63,* 83–89. http://dx.doi.org/10.1080/87567555.2015.1007913

Troisi, J. D., Leder, S., Stiegler-Balfour, J. J., Fleck, B. K. B., & Good, J. J. (2015). Effective teaching outcomes associated with the mentorship of early career psychologists. *Teaching of Psychology, 42,*

242–247. http://dx.doi.org/10.1177/009862 8315587623

Vansteenkiste, M., Simons, J., Lens, W., Sheldon, K. M., & Deci, E. L. (2004). Motivating learning, performance, and persistence: The synergistic effects of intrinsic goal contents and autonomy-supportive contexts. *Journal of Personality and Social Psychology, 87,* 246–260. http://dx.doi.org/10.1037/0022-3514.87.2.246

Warren, C. S. (2006). Incorporating multiculturalism into undergraduate psychology courses: Three simple active learning activities. *Teaching of Psychology, 33,* 105–109. http://dx.doi.org/10.1207/s15328023top3302_5

Williams, G. C., & Deci, E. L. (1996). Internalization of biopsychosocial values by medical students: A test of self-determination theory. *Journal of Personality and Social Psychology, 70,* 767–779. http://dx.doi.org/10.1037/0022-3514.70.4.767

Williams, G. C., Saizow, R., Ross, L., & Deci, E. L. (1997). Motivation underlying career choice for internal medicine and surgery. *Social Science & Medicine, 45,* 1705–1713. http://dx.doi.org/10.1016/S0277-9536(97)00103-2

Wilson-Doenges, G., & Gurung, R. A. R. (2013). Benchmarks for scholarly investigations of teaching and learning. *Australian Journal of Psychology, 65,* 63–70. http://dx.doi.org/10.1111/ajpy.12011

Wong, M. M.-H. (2000). The relations among causality orientation, academic experience, academic performance, and academic commitment. *Personality and Social Psychology Bulletin, 26,* 315–326. http://dx.doi.org/10.1177/0146167200265005