4

This chapter suggests solutions to common practical problems in designing SoTL studies. In addition, the advantages and disadvantages of different types of designs are discussed.

Designing SoTL Studies—Part II: Practicality

Robert A. Bartsch

For a study to succeed, a design has to be valid, practical, and ethical. The previous chapter examined validity and Chapter 6 examines ethics as it relates to the Scholarship of Teaching and Learning (SoTL; see also Gurung and Schwartz 2009; Landrum and McCarthy 2012). This chapter focuses on creating practical designs. Much of SoTL is field research. The studies are done in the student's natural environment, that is, the classroom. SoTL research can be tricky because designs may not be practical due to limited numbers of students, lack of time, or inability to have more than one condition. Other designs may be practical but are not ethical. For example, I could give half my class a study guide for an exam and nothing to the other half. The study could be a practical and valid test of the effectiveness of the study guide, but ethically challenged because some students and not others may receive a benefit (see also Swenson and McCarthy 2012). Given concerns of validity, practicality, and ethics, each study has advantages and disadvantages, and researchers have to determine the advantages they need and the disadvantages they can allow.

In the next section, I discuss some common practical problems and potential solutions. Then I discuss some common designs that can be used in SoTL research. As in the previous chapter, I focus on classroom activities, which I call *treatments* (i.e., the independent variables) and the measurement of their success, which I refer to as *assessments* (i.e., the dependent variables). Also, I continue to refer to instructors doing SoTL work as researchers.

Common Practical Problems

Many times faculty are interested in SoTL research, but they do not see how it can be done. In this section, I examine some common concerns new SoTL researchers have.

I Have to Measure Everything. Sometimes researchers get too ambitious and they measure many, many constructs. A researcher may want to determine everything about student-teacher interactions in a single study. The researcher bravely develops measures for student personality, student academic history, teaching style, students' preferred learning style, classroom characteristics, student learning, students' expected learning, students' perceived learning, and on and on. The worry is that an important construct will be missed which explains the entire area, and the strategy becomes measure everything and sort it out later. Even if every individual construct is measured well, too many constructs can cause problems with participant fatigue. Students are more likely to not think about the questions and respond with the first acceptable option and not the best answer (i.e., satisficing responses; Krosnick 1999). Additionally, too many measures make analyses much more complicated. Too often researchers create a mess of data and a headache rather than useful information. In these instances, a closer look at the literature and/or planning multiple studies can help researchers focus their study on only a few constructs.

I Do Not Have Many Students. One common limitation researchers have is that they do not have many students. Researchers may have only fifteen students in a course and may not teach the course again in the near future. A problem of testing few students is low statistical power (Wilson-Doenges and Gurung 2013). Statistical power is the ability to detect a significant finding that exists (Shadish, Cook, and Campbell 2002). In other words, low power indicates it is very likely one will not find any significant statistical effect. Researchers have only a 50 percent chance of finding a significant difference for a large effect with thirteen students in a treatment and thirteen students in a control condition (Gall, Gall, and Borg 2007). For a 70 percent chance of detecting a difference, researchers need forty total students. If it is a medium-sized effect, researchers need sixty-four students for only a 50/50 chance at finding an effect.

Power can be increased using several different methods. If one cannot increase the sample size, one can improve power by having a pre-test/post-test or other form of within-participants design (Shadish, Cook, and Campbell 2002). Researchers can detect differences more easily by seeing how students themselves change in these types of designs than other designs which separate students into treatment and control conditions.

I Only Have a Single Class. Another possible problem is the researcher only has a single class. Some designs can be used with a single group of students; however, the researcher will have greater flexibility in study design if the class can be divided. Dividing the class can cause ethical problems because only some students in the class will receive a treatment while others will be in a control condition.

LoSchiavo, Shatz, and Poling (2008) provide several ideas for how to practically split a class into groups. If the class has a web component, treatments can be given through the class course shell to only some students.

If the class splits into recitation sections, the treatment could be given during some recitation sections but not others. If neither of these is possible, the usual class time can be split such that some students arrive during the first part of class and receive the control condition and other students arrive during the second part of class and receive the treatment condition. Smith (2008) also suggests this method but goes further and states to minimize ethical concerns, students in the control condition could receive the treatment at the beginning of the next class, and the class could reunite as a whole later in the period. Researchers can also split the class and move them to separate rooms. Researchers could provide the treatment in one room and the control condition in the other (Bartsch, Case, and Meerman 2012). Finally, under some conditions, a researcher could give both the treatment and control conditions simultaneously to students in the same classroom. For example, to study the effect of question order in exams, the treatment condition may have exam questions in the order they were covered in class. The questions in the control condition may be in random order. These exams could be administered at the same time. In short, researchers have many possible options for splitting a single class into multiple groups.

Random Assignment Sounds Great, But How Can I Do It? Random assignment of students to treatment and control conditions creates high internal validity for the study (Shadish, Cook, and Campbell 2002). Unfortunately, random assignment does not occur naturally. Having two separate classes, whether they are from the same or different semesters, is not random assignment. Except in unusual situations, students would not have had an equal chance of being in each class. Students likely self-selected which course to take based on time, day of class, availability of other classes, or whether a friend was in the class. In other words, the students in each section may differ, and these differences could be confounds. For random assignment to occur with two separate classes, students have to be randomly sorted into the two classes (e.g., students with odd-numbered student IDs are in one section, even in the other), which is uncommon in higher education. Even if random assignment occurs, classes always differ even if they are taught similarly, creating confounds (Grauerholz and Main 2013).

Random assignment can be achieved inside a single classroom more easily. In all options mentioned in the last section for splitting classes, researchers can use random assignment to assign students to the treatment and control conditions. The class can be split by choosing numbers out of a hat, flipping a coin, using a random number table, and so on (Shadish, Cook, and Campbell 2002). Even assigning students by counting "1–2–1–2 . . . " achieves random assignment.

I Want to Detect a Subtle Effect. Another issue to consider is whether the treatment is large enough to have an effect on the assessment. For example, using an electronic classroom response system or "clickers" one time in class or even once each week for several weeks may not cause enough of a change in students to affect the assessment. In contrast, extensive use of clickers during the semester may demonstrate a detectable effect. Larger effects have greater statistical power (Shadish, Cook, and Campbell 2002). Higher power through testing large effects can equalize other factors that decrease power such as a small sample size. For example, to have a 70 percent chance of detecting a statistically significant correlation, researchers need 23 students for a large effect, 66 students for a medium-sized effect, and 616 students for a small effect (Gall, Gall, and Borg 2007). I recommend SoTL researchers first investigate larger effects. Larger effects are easier to detect, provide instructors evidence for treatments that cause bigger changes, and provide a better foundation for future studies to explore more nuanced effects.

I Cannot Do a Valid, Ethical Study in the Classroom. Occasionally, SoTL researchers try to create a valid, practical, and ethical field study but cannot produce a design with all three qualities. In this case, one strategy for creating a valid design is to move the study from the classroom to the laboratory, an environment where the researcher controls all the factors and does not have to worry about the ethics of teaching a class at the same time (see Bartsch and Murphy 2011, and Mayer and Johnson 2010 for examples). In the laboratory, a researcher has control to randomly assign participants who act as students (and generally are students) to conditions. Additionally, the increased control in lab studies allows researchers to more easily separate out important factors. Laboratory designs have some disadvantages. First, researchers need additional time and resources including recruiting students and finding a place to conduct the study. The second disadvantage is one of realism. Because students know the study is not for a grade, they may not try as hard as in a normal classroom situation. Lab studies complement the difficulties inherent in field studies and are a good way to begin exploration of a SoTL question.

Designs for Classroom Studies

In this section, I discuss many types of designs researchers can use in SoTL. For each design I provide a symbolic description. I have also included suggestions on when and when not to use each design. Although I present names of each design, I caution that different researchers and reference books may use different names. Of course, the name is not as important as the properties of the design. These designs are some of the more common examples. Other designs are often based on combining different parts of ones detailed here.

Simple Correlation. Using this design (Table 4.1), researchers examine the relationships between two or more variables. For example, Baker (2010) investigated learning in online classes and correlated several variables including instructor presence, instructor immediacy, student affective learning, student motivation, and student cognitive learning. In this type of study, researchers simply measure each construct. These measurements can

Table 4.1. Simple Correlation

	Assessment 1 ↔ Assessment 2
Use If	• Have single group of students that cannot be divided • Have only one session to collect data
Do Not Use If	• Want to make statement about causality • Have low number of students
Additional Options	• Correlate many variables at same time

Table 4.2. One-Group Post-Test Only Design

	Treatment → Assessment
Use If	• Desired focus is on describing treatment and not assessment • Cannot have pre-test or control group • Have single group of students that cannot be divided
Do Not Use If	• Want to make statement about causality • Want to make comparison to another group

be collected at the same or different times. However, in these designs, it is nearly impossible to establish that one variable caused changes in another variable (Shadish, Cook, and Campbell 2002). Because all students are measured on all assessments, this design can be used when one has a single group of students.

One-Group Post-Test Only Design. In this design (Table 4.2), the researcher exposes students to a treatment and later assesses students. For example, a study may describe the novel instruction over an entire semester, and the focus of the study is about the instruction. The researchers may only collect data, such as student learning or student attitudes, to demonstrate that after taking the class, the students have a certain level of knowledge, skill, or attitudes. This conclusion of obtaining a certain level is very different from saying that a treatment caused a change or an improvement. Without a comparison group, one cannot determine if a treatment had any effect. This design is generally not recommended, and in attempting to publish the study, the design would likely be viewed as flawed.

Two-Group Post-Test Only Design. This design is the simplest of the two-group designs. In the two-group post-test only design (Table 4.3), one group receives the treatment and another does not, and then both are measured on the same assessment. For example, one set of researchers explored whether an active learning class with clicker use, student-student discussions, small group tasks, and instructor feedback on these activities would lead to more student learning and engagement than a traditional lecture with some clicker questions (Deslauriers, Schelew, and Wieman 2011). They used two separate sections of the same course and used the same exam.

Table 4.3. Two-Group Post-Test Only Design

Some Students: Treatment → Assessment *Other Students: No Treatment → Assessment*	
Use If	• Concerned about carryover effects • Concerned about testing and instrumentation effects • Have multiple groups • Have only one session to collect data
Do Not Use If	• Have low number of students • Groups are very different • Have different assessments for each condition
Additional Options	• Use random assignment to improve internal validity • Add post-test to assess long-term change • Add additional conditions • Use covariates to improve internal validity and power

This design has no difficulty with carryover, testing, or instrumentation effects because each student has only one treatment and is assessed one time. However, as mentioned in the last chapter, the main problem is selection bias (Shadish, Cook, and Campbell 2002). The greater the difference between the two groups, the more arguments can be made that these differences between students caused any changes on the assessment and not the treatment. If random assignment is used, then researchers can avoid questions about selection bias because random assignment should equalize groups on all characteristics except for the treatment (Shadish, Cook, and Campbell 2002). A second problem with the design has to do with the relatively low statistical power of a comparison between groups (Shadish, Cook, and Campbell 2002). If class sizes are small, significant differences are difficult to detect.

Researchers may decide to have more than one treatment and one control condition. For example, Chang, Sung, and Chen (2001) studied the effects of concept mapping on learning. They had three conditions. Students either constructed a concept map by themselves, using a computer without hints, or using a computer with hints. In this case, the researchers not only can test the difference between not using and using a computer, but also how to best use the computer. Also, researchers may have more than two conditions because the treatment occurs in a range. As an example, researchers investigated how instructor self-disclosure on Facebook affected students' impression of the class instructor (Mazer, Murphy, and Simonds 2007). The researchers created high, medium, and low self-disclosure levels for the instructor. Using three levels allowed researchers to see the effect at the extremes and a more moderate value. Of course, studies with more than two conditions need more students. Researchers should have a minimum of ten students in each condition (Wilson-Doenges and Gurung 2013).

One-Group Pre-Test/Post-Test Design. With this design (Table 4.4), the researcher measures students before the treatment and then after

Table 4.4. One-Group Pre-Test/Post-Test Design

	Assessment → Treatment → Assessment
Use If	• Have low number of students • Have single group of students that cannot be divided • Cannot have control condition
Do Not Use If	• Items other than treatment occur between assessments • First assessment, by itself, affects second assessment • Students are likely to change between assessments with no treatment
Additional Options	• Add post-test to assess long-term change • Use alternative measures to minimize testing and instrumentation effects

Table 4.5. Two-Group Pre-Test/Post-Test Design

	Some Students: Assessment → Treatment → Assessment *Other Students: Assessment → No Treatment → Assessment*
Use If	• Have multiple groups
Do Not Use If	• Have single group of students that cannot be divided
Additional Options	• Use random assignment to improve internal validity • Add post-test to assess long-term change • Use alternative measures to minimize testing and instrumentation effects • Add additional conditions • Use covariates to improve internal validity and power

the treatment to determine any changes. For example, Bridges et al. (1998) wanted students to learn more quantitative reasoning in a non-method sociology course. After adding material to their class, they measured students at the beginning and end of the semester on quantitative reasoning.

Because it has a single group and compares students to themselves, the design is very useful with low numbers of students. This design has some disadvantages. Students or the environment may change naturally between the two assessments. This design cannot separate any natural changes from changes caused by the treatment. Additionally, the first assessment may affect the response on the second assessment (i.e., testing effect; Shadish, Cook, and Campbell 2002), especially if the questions are the same and the assessments are close together. Unfortunately, having different questions at the pre- and post-tests can lead to an instrumentation effect (Shadish, Cook, and Campbell 2002). These issues can be minimized using alternative measures (Bartsch, Bittner, and Moreno 2008).

Two-Group Pre-Test/Post-Test Design. This design (Table 4.5) has a treatment and control condition. Both groups are assessed before and after the treatment. As an example, Williams (2005) examined whether

Table 4.6. Within-Participants Design

Treatment 1 → Assessment → Treatment 2 → Assessment → [Continues]	
Use If	• Have low number of students • Have single group of students that cannot be divided
Do Not Use If	• Early treatments affect later treatments • Early assessments affect later assessments
Additional Options	• Add additional treatments • Counterbalance conditions to improve internal validity • Include pre-test to assess students before any treatment

participation in study abroad programs increased students' intercultural communication skills. Williams measured at the beginning and end of the semester students who participated and did not participate in a study abroad program.

With the pre-test, equalizing student differences between conditions becomes less important because the researcher has a baseline on each person for comparison. Of course, whenever possible researchers should randomly assign students to the treatment and control conditions.

The two-group pre-test/post-test design does not have major disadvantages, but because the design has multiple assessments, researchers should check on any testing and/or instrumentation effects. These can be minimized with alternative measures (Bartsch, Bittner, and Moreno 2008). Given the nature of field research and the difficulty of having random assignment, this design is often a good balance between validity and practicality.

Within-Participants Design. In this design (Table 4.6), each student is in each condition, and each student is assessed after each condition. Sometimes the study has a treatment and a control condition, or may have two similar treatment conditions. The study can also have more than two conditions. For example, one study examined the effectiveness of PowerPoint presentations (Bartsch and Cobern 2003). In this study, researchers rotated between three conditions: overhead transparencies, plain PowerPoint slides, and PowerPoint slides with pictures, graphs, transitions, and sound effects. Conditions rotated each week, and each week researchers quizzed the students.

The strength of within-participants designs is that researchers do not need as many students in their studies. Counterbalancing conditions is recommended to improve internal validity. With counterbalancing, different students take treatments in different orders to minimize carryover effects (Dunn 2009). Of course, to counterbalance, students have to be split into different groups.

Crossover Design. In this design (Table 4.7; Shadish, Cook, and Campbell 2002), students are split into two groups. One group receives the treatment, both groups are assessed, then the other group receives the treatment, and both groups are assessed again. The crossover design is a

Table 4.7. Crossover Design

	Some Students: Treatment → Assessment → No Treatment → Assessment *Other Students: No Treatment → Assessment → Treatment → Assessment*
Use If	• Have low number of students • Have multiple groups
Do Not Use If	• First assessment, by itself, affects second assessment • Have single group of students that cannot be divided
Additional Options	• Include pre-test to assess students before any treatment • Add post-test to examine long-term change • Use random assignment to improve internal validity • Use alternative measures to minimize testing and instrumentation effects

Table 4.8. Interrupted Time-Series Design

Multiple Assessments → Treatment → Multiple Assessments	
Use If	• Have low number of students • Have single group of students that cannot be divided • Want to determine long-term effects
Do Not Use If	• Have only one session to collect data • Early assessments affect later assessments
Additional Options	• Add control condition to improve internal validity • Add additional treatment condition, with treatment at different time to improve internal validity

counterbalanced within-participants design with two conditions. For example, Ocker and Yaverbaum (1999) wanted to determine any differences in performance and preference on a case study between face-to-face groups and groups using asynchronous computer collaboration. They used two case studies. Some student groups did the first case study face-to-face and the second asynchronously on the computer. Other student groups did the first case study on the computer and the second face-to-face.

In crossover designs, researchers need to make sure the first assessment does not by itself cause changes on the second assessment. This potential problem can be minimized by using related assessments or by alternative measures. In Ocker and Yaverbaum's research it would have been meaningless to assess the same case study under both conditions, and consequently they used two separate cases.

Interrupted Time-Series Design. In this design (Table 4.8; Shadish, Cook, and Campbell 2002), multiple pre-tests occur before and multiple post-tests occur after the treatment. For example, suppose researchers want to assess students' self-efficacy across time. They measure self-efficacy every two weeks. Halfway through the semester, the researchers stage an intervention to change self-efficacy. This time-series design can determine the stability of scores before treatment (i.e., the intervention) and can examine the effect of the treatment over a longer period of time. This design minimizes the problem with a one sample pre-test/post-test design of students naturally changing.

Researchers can also include a control condition. The same assessments occur at the same time but students in the control condition do not receive the treatment. The control condition strengthens the internal validity of the study. Similarly, researchers could also have another condition but that condition receives the treatment at a different time. In this design, we expect similar changes to occur for both groups after they receive the treatment condition. Because treatments occur at different times, the predicted change should occur at different times. This design rules out many possible confounds enhancing the internal validity of the study. These designs work well without random assignment. Of course, another group of students needs to be available.

More Complex Designs

As SoTL research becomes more sophisticated, we will begin to move away from simpler research questions such as whether this treatment causes a change in assessment. Rather researchers will build on the simple statement to more complex questions. This final section looks at some of these designs.

Use Multiple Treatments to Investigate Interactions. So far in the chapter, I discussed designs with only one treatment or independent variable. However, many research designs have more than one treatment. For example, Chesbro (2003) manipulated both nonverbal immediacy of an instructor and instructor clarity to determine how students respond to both student learning and student affect. Nonverbal immediacy and clarity are two separate treatments. Each one may have an effect and an interaction may occur between the two treatments. Think of these interactions like medical drug interactions. Sometimes combining different drugs causes very different results than just adding their individual effects. Only by testing multiple treatments at once can researchers investigate how the treatments interact with each other.

Use Moderators to Determine When Treatment Has Effect. Instead of asking whether a treatment works, researchers may ask under what conditions or when does the treatment change the assessment. The moderator alters the effect of the treatment on the assessment (Baron and Kenny 1986). Cole, Field, and Harris (2004) investigated whether psychological hardiness moderated the effect of learning motivation on reactions to classroom experience. In other words, they asked whether students who are low on psychological hardiness have a different motivation-experience relationship than students high on psychological hardiness. In another example, I may be interested in whether clickers work better on first-generation college students or non-first-generation college students. In this case, whether a student is first-generation is the moderating variable, and the relationship between clickers and student success may differ between first-generation and non-first-generation students. These questions allow researchers to probe under what conditions these effects exist.

A moderating variable is statistically equivalent to a second treatment described in the previous section. The difference is the label and context. Oftentimes using multiple treatments, the researcher is interested in all treatments individually and their interaction. With moderators, researchers are more interested in the effect the treatment has on the assessment and how the moderating variable influences that effect. Typically, researchers do not care about the effect of the moderator by itself on the assessment.

Use Mediators to Investigate How Treatment Has Effect. Another more complex design is a mediational analysis. The main question asked in this type of analysis is how the treatment causes change in the assessment. Mediators delve into the process of what happens with students and "account for the relation" between the treatment and assessment (Baron and Kenny 1986, 1176). For example, Elliot, McGregor, and Gable (1999) wanted to investigate more deeply the relationship between motivational goals and exam scores. They found persistence of effort mediated the relationship between performance approach goals and exam scores, and disorganization mediated the relationship between performance avoidance goals and exam scores. Therefore, according to their analysis, a performance approach goal leads to more persistence which then leads to higher exam scores and a performance avoidance goal leads to more disorganization which leads to lower exam scores. These mediators suggest the mechanism behind the overall motivational goal-exam score relationship.

Conclusion

SoTL research can be difficult given its many practical limitations. Table 4.9 summarizes the different types of designs presented and which designs should be used with small samples or single classes. The table also describes which designs allow researchers to complete the study in a single session, whether the design is set to detect long-term effects, and if the design supports causal claims. The suggestions provided in the table and this chapter hopefully will help future researchers create practical designs to test their research questions.

Each design has advantages and disadvantages. Although there may be better and worse ways to design and conduct a study, there will never be *one* clear right way. There is no single ideal study that eliminates all potential problems and all alternative hypotheses. There is no one study that answers

Table 4.9. Summary of When and When Not to Use Designs

Design	*Use with Small Samples*	*Use with Single Class*	*Can Complete in One Session*	*Can Determine Long-term Effects*	*Can Determine Causal Relationships*	*Main Weaknesses*
Simple correlation	No	Yes	Yes	No	No	Cannot determine causality
One-group post-test only	Yes	Yes	Yes	No	No	Cannot make comparisons; cannot determine causality
Two-group post-test only	No	If class can be divided	Yes	If add long-term post-test	Yes with random assignment; typically no without	Selection bias
One-group pre-test/post-test	Yes	Yes	If pre- and post-test in same session	If add long-term post-test	Depends on context	Testing and instrumentation effects; confounds between assessments
Two-group pre-test/post-test	Maybe	If class can be divided	If pre- and post-test in same session	If add long-term post-test	Typically yes, better with random assignment	Various minor issues depending on context
Within participants	Yes	Yes	If all treatments and assessments in same session	No	Typically yes if counterbalanced	Carryover, testing, and instrumentation effects
Crossover	Yes	If class can be divided	If both treatments and assessments in same session	If add long-term post-test	Typically yes	Testing and instrumentation effects
Interrupted time series	Yes	Yes	No	Yes	Typically yes	Testing and instrumentation effects

all questions. The strength of science, including SoTL research, does not lie in one individual study, but rather that a large number of studies from a large number of researchers who together push the boundaries of what we know.

References

Baker, C. 2010. "The Impact of Instructor Immediacy and Presence for Online Student Affective Learning, Cognition, and Motivation." *Journal of Educators Online* 7 (1): 1–30.

Baron, R. M., and D. A. Kenny. 1986. "The Moderator–Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations." *Journal of Personality and Social Psychology* 51: 1173–1182. doi:10.1037/0022-3514.51.6.1173.

Bartsch, R. A., W. M. E. Bittner, and J. E. Moreno. 2008. "A Design to Improve Internal Validity of Assessments of Teaching Demonstrations." *Teaching of Psychology* 35: 357–359. doi:10.1080/00986280802373809.

Bartsch, R. A., K. A. Case, and H. Meerman. 2012. "Increasing Academic Self-Efficacy in Statistics with a Live Vicarious Experience Presentation." *Teaching of Psychology* 39: 133–136. doi:10.1177/0098628312437699.

Bartsch, R. A., and K. Cobern. 2003. "Effectiveness of PowerPoint Presentations in Lectures." *Computers and Education* 41: 77–86.

Bartsch, R. A., and W. Murphy. 2011. "Examining the Effects of an Electronic Classroom Response System on Student Engagement and Performance." *Journal of Educational Computing Research* 44 (1): 25–33. doi:10.2190/EC.44.1.b.

Bridges, G. S., G. M. Gillmore, J. L. Pershing, and K. A. Bates. 1998. "Teaching Quantitative Research Methods: A Quasi-Experimental Analysis." *Teaching Sociology* 26: 14–28.

Chang, K. E., Y. T. Sung, and S. F. Chen. 2001. "Learning through Computer-Based Concept Mapping with Scaffolding Aid." *Journal of Computer Assisted Learning* 17: 21–33.

Chesbro, J. L. 2003. "Effects of Teacher Clarity and Nonverbal Immediacy on Student Learning, Receiver Apprehension, and Affect." *Communication Education* 52: 135–147.

Cole, M. S., H. S. Field, and S. G. Harris. 2004. "Student Learning Motivation and Psychological Hardiness: Interactive Effects on Students' Reactions to a Management Class." *Academy of Management Learning and Education* 3: 64–85.

Deslauriers, L., E. Schelew, and C. Wieman. 2011. "Improved Learning in a Large-Enrollment Physics Class." *Science* 332: 862–864. doi:10.1126/science.1201783.

Dunn, D. S. 2009. *Research Methods for Social Psychology*. Malden, MA: Wiley-Blackwell.

Elliot, A. J., H. A. McGregor, and S. Gable. 1999. "Achievement Goals, Study Strategies, and Exam Performance: A Mediational Analysis." *Journal of Educational Psychology* 91: 549–563.

Gall, M. D., J. P. Gall, and W. R. Borg. 2007. *Educational Research: An Introduction*, 8th ed. Boston: Pearson.

Grauerholz, L., and E. Main. 2013. "Fallacies of SoTL: Rethinking How We Conduct Our Research." In *The Scholarship of Teaching and Learning in and across the Disciplines*, edited by K. McKinney, 152–168. Bloomington: Indiana University Press.

Gurung, R. A. R., and B. M. Schwartz. 2009. *Optimizing Teaching and Learning: Practicing Pedagogical Approach*. Chichester, West Sussex, UK: Wiley-Blackwell.

Krosnick, J. A. 1999. "Survey Research." *Annual Review of Psychology* 50: 537–567.

Landrum, R. E., and M. A. McCarthy. 2012. *Teaching Ethically: Challenges and Opportunities*. Washington, DC: American Psychological Association.

LoSchiavo, F. M., M. A. Shatz, and M. A. Poling. 2008. "Strengthening the Scholarship of Teaching and Learning via Experimentation." *Teaching of Psychology* 35: 301–304. doi:10.1080/00986280802377164.

Mayer, R. E., and C. I. Johnson. 2010. "Adding Instructional Features that Promote Learning in a Game-Like Environment." *Journal of Educational Computing Research* 42: 241–265. doi:10.2190/EC.42.3.a.

Mazer, J. P., R. E. Murphy, and C. J. Simonds. 2007. "I'll See You on 'Facebook': The Effects of Computer-Mediated Teacher Self-Disclosure on Student Motivation, Affective Learning, and Classroom Climate." *Communication Education* 56 (1): 1–17. doi:10.1080/03634520601009710.

Ocker, R. J., and G. J. Yaverbaum. 1999. "Asynchronous Computer-Mediated Communication versus Face-to-Face Collaboration: Results on Student Learning, Quality and Satisfaction." *Group Decision and Negotiation* 8: 427–440.

Shadish, W. R., T. D. Cook, and D. T. Campbell. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin.

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. doi:10.1080/00986280802418711.

Swenson, E. V., and M. A. McCarthy. 2012. "Ethically Conducting the Scholarship of Teaching and Learning Research." In *Teaching Ethically: Challenges and Opportunities*, edited by R. Landrum and M. McCarthy, 21–29. Washington, DC: American Psychological Association.

Williams, T. R. 2005. "The Impact of Study Abroad on Students' Intercultural Communication Skills: Adaptability and Sensitivity." *Journal of Studies in International Education* 9: 356–371. doi:10.1177/1028315305277681.

Wilson-Doenges, G., and R. A. R. Gurung. 2013. "Benchmarks for Scholarly Investigations of Teaching and Learning." *Australian Journal of Psychology* 65: 63–70. doi:10.1111/ajpy.12011.

ROBERT A. BARTSCH is an associate professor of psychology at the University of Houston-Clear Lake.