ABSTRACT

Studying the effectiveness of educational interventions is centrally important to building the science of nursing education. Yet, the design most commonly used in the study of nursing education interventions—the single-group, pre- and posttest design—provides limited evidence to support claims of intervention effectiveness. In this Methodology Corner installment, the limitations of the single-group, pre- and posttest design are outlined and a review of the requirements for establishing stronger arguments for causality is presented. To overcome the limitations of single-group, pre- and posttest designs, nursing education researchers are encouraged to employ study designs and procedures that can significantly strengthen researchers’ claims of intervention effectiveness. [J Nurs Educ. 2018;57(2):69-71.]

Among a variety of critical decisions researchers must make when designing a study, those that impact the study’s internal and external validity, and thus its usefulness to nurse educators, are among the most important. Although discussions of the fundamental importance of the validity and reliability of study measures are numerous (Ironside & Spurlock, 2014; Sepucha et al., 2014; Spurlock, 2017), in this Methodology Corner installment, we focus on the issue of basic study design and, specifically, on the use of single-group, pre- and posttest designs for evaluating educational interventions. Calls to move on from this design are not new (Campbell & Stanley, 1963); yet, the design and its accompanying detriments persist—and even dominate the nursing education research literature to this day.

Single-Group Designs in Nursing Education Research

In an immensely insightful study, which cries out for replication with more recent, published research, Yucha, Schneider, Smyer, Kowalski, and Stowers (2011) evaluated the methodological quality of 133 nursing education research reports published between 2006 and 2007. Using an established educational research quality rating tool, the Medical Education Research Study Quality Instrument (MERSQI) (Reed et al., 2007), Yucha et al. found that 74.4% \((n = 99)\) of the reviewed studies used single-group designs, which receive the lowest score on the MERSQI. Nonrandomized designs using two or more groups were used in 21.8% of cases, and only five studies (3.8%) used a randomized controlled trial design. Reporting similar findings in a study focusing on nursing school deans’ \((N = 12)\) perspectives on the state of the science of nursing education research, Broome, Ironside, and McNelis (2012) reported that when asked about the types of research designs used in their schools, 90% of deans indicated that none of the nursing education research being conducted in their schools was experimental in nature. Concerns about using single-site, descriptive, single-group design have been echoed in recent years in multiple journals publishing nursing education research (Ironside & Spurlock, 2012; Morton, 2017). Recently, Morton (2017) articulated the issue with exacting precision, noting, “A commonly used methodology is a pretest, posttest with a single-group of learners—Reviewers find these types of studies to have minimal relevance to a larger audience and contribute little to the generation of new and useful knowledge” (pp. 311-312).


**Limitations of the Single-Group Design**

In its most common form, the single-group design involves pre- and postintervention testing of variables such as knowledge, attitudes, satisfaction, or skills in a single-group of subjects; interventions are typically educational or behavioral in nature. The primary weakness of this design is the numerous threats to internal validity, which Knapp (2016) so aptly reminds us impacts primarily the interpretation of causal inference. The most commonly identified threats to internal validity, learned by all nursing students in basic through advanced nursing research courses, were described long ago by Campbell and Stanley (1963) as history, maturation, testing, instrumentation, and statistical regression. Other threats have since been added, but for this discussion, this longstanding list will be the focus. Each of these threats involve events and experiences that can happen to subjects between the pre- and the postintervention period and that interfere with the researcher’s ability to conclude that any changes in the scores of outcome variables can be attributed only to the intervention and not to other factors. A more thorough discussion of threats to internal validity can be found in any nursing research textbook. Other recommended sources include those by Knapp (2016) and Marsden and Torgerson (2012).

To highlight how various threats to validity come to bear on a study’s internal validity, a commonly encountered study archetype is described in this article. Interested in further exploring an anecdotal observation from teaching, nursing education researchers measure students on one or more variables during the preintervention period and expose the students to an intervention over the course of hours (such as in the case of 1-day workshops), weeks (as with short remediation programs), or months (such as with semester-long courses). Students are then measured again postintervention to determine intervention effectiveness. The threats to internal validity in this scenario are innumerable. The testing threat manifests when students’ scores improve from pre- to postintervention simply from exposure to the study instruments during the pretest because they were exposed to the study instruments during the pretest. They may score higher on the posttest due simply due to familiarity and having more time to reflect on the instruments. If the intervention period is longer than a few minutes and students are not under constant observation, technology can play a role too. Internet-connected mobile devices provide a quick way to learn something outside of the intervention itself! The threat of maturation manifests generally when intervention periods are long. For example, improvements seen on measures of students’ coping skills between the pretest and posttest could be attributed to their participation in a year-long online coping skill-building group, or perhaps the students accumulated additional coping skills naturally in the face of adversities encountered in their personal lives during that year. How can we know?

Finally, the threat of statistical regression, sometimes also called *regression toward the mean*, occurs when individuals who scored at the margins of the scoring distribution (either high or low) during the pretest tend to move toward the middle (i.e., the mean) of the score distribution at the posttest. This effect is sometimes explained in terms of luck: pretest low and high scorers tend to be more or less lucky in their scoring, respectively, in the posttest period. The statistical regression effect is more pronounced when measures with low reliability are used, as low instrument reliability introduces increased levels of measurement error. Marsden and Torgerson (2012) provided a complete and accessible explanation of the statistical regression threat.

The effect of these threats can be seen in substantial evidence from the social sciences, which suggest that single-group, pre- and posttest study effect sizes are overestimated by an average of 61% compared with those studies using a control group (Lipsey & Wilson, 1993). In evaluating the quality of studies of Internet-based instruction in the health professions, Cook, Levinson, and Garside (2011) similarly found that effect sizes were 59% higher in single-group verses studies with two or more groups ($p = .013$). In a recent systematic review and meta-analysis of simulation-based training for cardiac auscultation skills (McKinney, Cook, Wood, & Hatala, 2012), single-group studies reported effect sizes for skill improvement that were four times higher than two-group studies (1.41 verses .35; respectively)!  

**Causality and Solutions to Single-Group Designs**

Threats to the internal validity of a study may seem like abstract, primarily theoretical challenges that require only a casual assessment before moving forward with study design and implementation. But, if we accept that threats to internal validity are actually threats that can abolish the researcher’s ability to establish a credible argument for causation, and thus render the findings of a study meaningless, then it becomes indisputably clear that minimizing threats to internal (and also external) validity should rise to the top of the checklist for designing studies in nursing education research.

The 19th century philosopher John Stuart Mills proposed rules for explaining causation that have been widely adopted and aptly explained in research texts such as those by Polit and Beck (2016) and the seminal education research text by Cohen, Manion, and Morrison (2017). To summarize here, there are three required conditions before researchers can claim that an effect was created by a cause. First, the cause must precede the effect in time. Successful performance of intravenous insertion cannot be attributed to teaching if the intravenous insertion occurred before the teaching. Second, differing levels of the cause must be associated in a theoretically plausible way with differing levels of the effect. So, if more time studying for an examination yields higher examination scores, then less time studying for an examination should yield lower examination scores. The third condition to establish causation, and perhaps the most challenging for researchers to establish, is that the effect observed cannot be explained by plausible confounding causes, or as they are more conventionally termed, *confounding variables*. If students are permitted to make use of their textbooks and electronic resources during examinations, then the impact of instruction on student learning cannot be evaluated because the effects of instruction and the effects of using reference sources during the examination are impossibly, inseparably confounded.

Understandably, not all potential confounding variables can be identified, measured, and accounted for, especially in descriptive, correlational research. But in intervention research, truly
experimental research designs provide substantial strength to researchers’ ability to claim causation when explaining intervention effects. The relatively simple solution to the limitations of single-group, pre- and posttest research designs is the addition of a control group to which subjects are randomly assigned, measured during the pre- and posttest period, but who do not receive the intervention under investigation. Although researchers often claim that having a control group presents ethical or practical challenges that cannot be overcome, the both literal and figurative mountains of this type of research in the social, educational, and clinical literatures suggest otherwise. With the addition of a control group, and in the context of educational research, this can be a group that receives standard instruction instead of no instruction, and researchers can immediately improve the strength of their arguments that any observed effects were likely created by the cause, and not confounders. Although there are a variety of other research designs, such as repeated measures, crossover, and time series designs, that may also reduce the impact of measured and unmeasured confounding variables, those methods are often truly difficult to implement in educational settings and are also beyond the scope of this discussion.

Conclusion

In this article, we have identified that when the goal is to test an educational intervention, nursing education research is dominated by single-group, pre- and posttest design studies. These studies have faced a long history of strong criticism for their susceptibility to numerous threats to internal validity, and thus, for the relatively weak empirical evidence these studies provide. In addition, effect sizes from these types of studies are overinflated when compared with more rigorous designs and thus present an unclear picture of the true effect of the interventions being tested. The primary weakness of single-group, pre- and posttest designs is that they cannot satisfy the third requirement for establishing causation in research that postintervention effects cannot be plausibly explained by additional, confounding variables. The only real solution to the challenges presented by single-group, pre- and posttest studies is to stop conducting these studies and move toward the use of more rigorous designs, such as those that use randomization and control groups. Movement in this direction is long overdue but essential for advancing the science of nursing education.

References


