Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal.

Volume 5, Number 14, November, 1997 ISSN=1531-7714

True and Quasi-Experimental Designs.

Barry Gribbons
National Center for Research on Evaluation, Standards, and Student Testing

Joan Herman
National Center for Research on Evaluation, Standards, and Student Testing

Experimental designs are especially useful in addressing evaluation questions about the effectiveness and impact of programs. Emphasizing the use of comparative data as context for interpreting findings, experimental designs increase our confidence that observed outcomes are the result of a given program or innovation instead of a function of extraneous variables or events. For example, experimental designs help us to answer such questions as the following: Would adopting a new integrated reading program improve student performance? Is TQM having a positive impact on student achievement and faculty satisfaction? Is the parent involvement program influencing parents' engagement in and satisfaction with schools? How is the school's professional development program influencing teacher's collegiality and classroom practice?

As one can see from the example questions above, designs specify from whom information is to be collected and when it is to be collected. Among the different types of experimental design, there are two general categories:

- true experimental design: This category of design includes more than one purposively created group, common measured outcome(s), and random assignment. Note that individual background variables such as sex and ethnicity do not satisfy this requirement since they cannot be purposively manipulated in this way.
- quasi-experimental design: This category of design is most frequently used when it is not feasible for the researcher to use random assignment.

This article describes the strengths and limitations of specific types of quasi-experimental and true experimental design.

QUASI-EXPERIMENTAL DESIGNS IN EVALUATION

As stated previously, quasi-experimental designs are commonly employed in the evaluation of educational programs when random assignment is not possible or practical. Although quasi-experimental designs need to be used commonly, they are subject to numerous interpretation problems. Frequently used types of quasi-experimental designs include the following:

Nonequivalent group, posttest only (Quasi-experimental).

The nonequivalent, posttest only design consists of administering an outcome measure to two groups or to a program/treatment group and a comparison. For example, one group of students might receive reading instruction using a whole language program while the other receives a phonetics-based program. After twelve weeks, a reading comprehension test can be administered to see which program was more effective.

A major problem with this design is that the two groups might not be necessarily the same before any instruction takes place and may differ in important ways that influence what reading progress they are able to make. For instance, if it is found that the students in the phonetics groups perform better, there is no way of determining if they are better prepared or better readers even before the program and/or whether other factors are influential to their growth.

Nonequivalent group, pretest-posttest.

The nonequivalent group, pretest-posttest design partially eliminates a major limitation of the nonequivalent group, posttest only design. At the start of the study, the researcher empirically assesses the differences in the two groups. Therefore, if the researcher finds that one group performs better than the other on the posttest, s/he can rule out initial differences (if the groups were in fact similar on the pretest) and normal development (e.g. resulting from typical home literacy practices or other instruction) as explanations for the differences.

Some problems still might result from students in the comparison group being incidentally exposed to the treatment

condition, being more motivated than students in the other group, having more motivated or involved parents, etc. Additional problems may result from discovering that the two groups do differ on the pretest measure. If groups differ at the onset of the study, any differences that occur in test scores at the conclusion are difficult to interpret.

Time series designs.

In time series designs, several assessments (or measurements) are obtained from the treatment group as well as from the control group. This occurs prior to and after the application of the treatment. The series of observations before and after can provide rich information about students' growth. Because measures at several points in time prior and subsequent to the program are likely to provide a more reliable picture of achievement, the time series design is sensitive to trends in performance. Thus, this design, especially if a comparison group of similar students is used, provides a strong picture of the outcomes of interest. Nevertheless, although to a lesser degree, limitations and problems of the nonequivalent group, pretest-posttest design still apply to this design.

TRUE EXPERIMENTAL DESIGNS

The strongest comparisons come from true experimental designs in which subjects (students, teachers, classrooms, schools, etc.) are randomly assigned to program and comparison groups. It is only through random assignment that evaluators can be assured that groups are truly comparable and that observed differences in outcomes are not the result of extraneous factors or pre-existing differences. For example, without random assignment, what inference can we draw from findings that students in reform classrooms outperformed students in non-reform classrooms if we suspect that the reform teachers were more qualified, innovative, and effective prior to the reform? Do we attribute the observed difference to the reform program or to pre-existing differences between groups? In the former case, the reform appears to be effective, likely worth the investment, and possibly justifying expansion; in the latter case, alternative inferences are warranted. There are several types of true experimental design:

Posttest Only, Control Group.

Posttest only, control group designs differ from previously discussed designs in that subjects are randomly assigned to one of the two groups. Given sufficient numbers of subjects, randomization helps to assure that the two groups (or conditions, raters, occasions, etc.) are comparable or equivalent in terms of characteristics which could affect any observed differences in posttest scores. Although a pretest can be used to assess or confirm whether the two groups were initially the same on the outcome of interest(as in pretest-posttest, control group designs), a pretest is likely unnecessary when randomization is used and large numbers of students and/or teachers are involved. With smaller samples, pretesting may be advisable to check on the equivalence of the groups.

Other Designs.

Some other general types of designs include counterbalanced and matched subjects (for a more detailed discussion of different designs see Campbell & Stanley, 1966). With counterbalanced designs, all groups participate in more than one randomly ordered treatment (and control) conditions. In matched designs, pairs of students matched on important characteristics (for example, pretest scores or demographic variables) are assigned to one of the two treatment conditions. These approaches are effective if randomization is employed.

Even experimental designs, however, can be problematic even when true experimental designs are employed (Cook & Campbell, 1979). One threat is that the control group can be inadvertently exposed to the program; such a threat also occurs when key aspects of the program also exist in the comparison group. Additionally, one of the conditions (groups), such as instructional programs may be perceived as more desirable than the other. If participants in the study learn of the other group, then important motivational differences (being demoralized or even trying harder to compensate) could impact the results. Differences in the quality with which a program or comparison treatment is implemented also can influence results (the teachers implementing one or the other have greater content or pedagogical knowledge). Still another threat to the validity of a design is differential participant mortality in the two groups.

LIMITATIONS OF TRUE EXPERIMENTAL DESIGN

Experimental designs also are limited by narrow range of evaluation purposes they address. When conducting an evaluation, the researcher certainly needs to develop adequate descriptions of programs, as they were intended as well as how they were realized in the specific setting. Also, the researcher frequently needs to provide timely, responsive feedback for purposes of program development or improvement. Although less common, access and equity issues within a critical theory framework may be important. Experimental designs do not address these facets of evaluation.

With complex educational programs, rarely can we control all the important variables which are likely to influence program outcomes, even with the best experimental design. Nor can the researcher necessarily be sure, without verification, that the implemented program was really different in important ways from the program of the comparison group(s), or that the implemented program (not other contemporaneous factors or events) produced the observed results. Being mindful of these issues, it is important for evaluators not to develop a false sense of security.

Finally, even when the purpose of the evaluation is to assess the impact of a program, logistical and feasibility issues

constrain experimental frameworks. Randomly assigning students in educational settings frequently is not realistic, especially when the different conditions are viewed as more or less desirable. This often leads the researcher to use quasi-experimental designs. Problems associated with the lack of randomization are exacerbated as the researcher begins to realize that the programs and settings are in fact dynamic, constantly changing, and almost always unstandardized.

RECOMMENDATIONS FOR EVALUATION

The primary factor which directs the evaluation design is the purpose for the evaluation. Restated, it is critical to consider the utility of any evaluation information. If the program's impact on participant outcomes is a key concern or if multiple programs (instructional strategies, or something else) are being considered and educators are looking for evidence to assess the relative effectiveness of each to inform decisions about which approach to select, then experimental designs are appropriate and necessary. Nonetheless, resulting information should be augmented by rich descriptions of programs and mechanisms need to be established which enable providing timely, responsive feedback (For a detailed discussion of other approaches to evaluation, see Lincoln & Guba, 1985; Patton, 1997, and Reinhart & Rallis, 1994).

In addition to using multiple evaluation methods, evaluators should be careful in collecting the right kinds of information when using experimental frameworks. Measures must be aligned with the program's goals or objectives. Additionally, it is often much more powerful to employ multiple measures. Triangulating several lines of evidence or measures in answering specific evaluation questions about program outcomes increases the reliability and credibility of results. Furthermore, when interpreting this evidence, it is often useful to use absolute standards of success in addition to relative comparisons.

The last recommendation is to always consider alternative explanations for any observed differences in outcome measures. If the treatment group outperforms the control group, consider a full range of plausible explanations in addition to the claim that the innovative practice is more effective. Program staff and participants can be very helpful in identifying these alternative explanations and evaluating the plausibility of each.

ADDITIONAL READING

Campbell, D.T. & Stanley, J.C. (1966). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally College Pub. Co.

Cook, T.D. & Campbell, D.T. (1979). *Quasi-experimentation: design and analysis issues for field settings*. Chicago: Rand McNally College Pub. Co.

Lincoln, Y.S. & Guba, E.G. (1985). *Naturalistic inquiry*. Beverly Hills: Sage Publications.

Patton, M.Q. (1997). *Utilization focused evaluation,* edition 3. Thousand Oaks, CA: Sage Publications.

Reinhart, C.S. & Rallis, S.F. (1994). *The qualitative-quantitative debate: New perspectives*. San Francisco: Jossey-Bass.

Descriptors: *Comparative Analysis; *Control Groups; Evaluation Methods; Evaluation Utilization; *Experiments; Measurement Techniques; *Pretests Posttests; *Quasiexperimental Design; Sampling; Selection

Citation: Gribbons, Barry & Herman, Joan (1997). True and quasi-experimental designs. *Practical Assessment, Research & Evaluation*, 5(14). Available online: http://PAREonline.net/getvn.asp?v=5&n=14.