Quasi-Experimental Design in Education

The following example illustrates what a quasi-experiment may look like in education:

A principal would like to know whether a recently implemented after-school program is positively impacting at-risk students’ academic achievement in math as measured by the Iowa Assessments. Since random assignment is impractical due to real-world constraints (the principal for ethical reasons cannot deny some at-risk students the opportunity of participating in an after-school program from which they may benefit) the principal decides to use a nonequivalent control group design with pretest and posttest measures (a type of quasi-experimental design) to compare the academic achievement of students who did not participate in the program (the control group) with those who participated in the program for 90 or more days.

Quasi-experiments have been used as far back as the 18th century and continue to be frequently utilized by researchers today for three primary reasons: (1) to meet the practical requirements of funding, school administrators, and ethics, (2) to evaluate the effectiveness of an intervention when the intervention has been implemented by educators prior to the evaluation procedure having been considered and (3) when researchers want to dedicate greater resources to issues of external and construct validity (Shadish, Cook, & Campbell, 2002).

Two limitations are generally associated with quasi-experiments. First, quasi-experiments do not allow researchers to determine the order by which variables occur. For instance, if a researcher were studying the influence of student motivation on academic achievement with a quasi-experiment, the researcher would never be able to determine whether student motivation causes academic achievement or whether consistent academic achievement positively impacts students’ motivation. Second, since quasi-experiments do not utilize random assignment of participants to groups, there is greater likelihood that extraneous variables (variables that are not the focus of the study that cause confusion to occur when researchers consider the relationships between variables being studied) (Slavin, 2007) may impact a study’s findings (Heiman, 1999). Researchers (e.g., Glazerman, Levy, & Myers, 2002; Heinsman & Shadish, 1996) concluded that quasi-experiments, when compared to true-experiments, often result in substantially different findings. Such differences were; however, considerably smaller for high quality quasi-experiments. Thus, quasi-experiments are based on creative design techniques to reduce the various threats that may cause a study’s findings to be invalid or unreliable (Green, 2006).

Design Elements of Quasi-Experiments

There are several elements that researchers must consider when designing a quasi-experiment. First, researchers must consider the issue of assignment – to which group individuals will be included during the experiment. While researchers do not control individuals’ group assignments in a quasi-experiment, researchers often need to consider whether they are going to allow participants to self-select the group to which they belong (Shadish, Cook, & Campbell, 2002). Research (e.g., Lipsey & Wilson, 1993) suggests that while random assignment (which does not occur in quasi-experiments) often leads to different study findings when compared to experiments that do use random assignment, differences in findings are even more pronounced in
experiments in which participants are allowed to self-select the group to which they belong (either the control or experiment group) (Heinsman & Shadish, 1996; Shadish, Matt, Navarro, & Phillips, 2000). Additional methods of group assignment can also be practiced by researchers including matching and stratifying which can be utilized to ensure that the control group and experiment group are as similar as possible at the start of the experiment. Matching and stratifying groups requires that researchers match participants in each group on as many characteristics as possible to ensure that control and experiment groups are as similar as possible before the treatment is introduced (Shadish et al., 2002).

Masking can also ensure that results from a quasi-experiment are as accurate as possible. Masking requires that researchers and participants remain blind (unknowing) of the group to which they have been assigned. This practice ensures that researchers and participants do not bias the results of the experiment by attempting to influence the outcomes of the experiment. Thus, not all random assignment practices in experiments are alike (2003).

Another issue researchers need to consider when designing a quasi-experiment is the way in which a particular outcome (dependent variable) will be measured at the end of the treatment period. A posttest is often utilized by researchers for this purpose (Slavin, 2007). One kind of posttest often used by researchers in education is a non-equivalent dependent variable. This particular posttest measures two constructs, one of which is expected to change after a treatment is enacted while the other is expected to remain consistent with previous results. For instance, in the example provided earlier of the principal wanting to study the impact of an after-school program on students’ academic achievement in math by using a non-equivalent dependent variable type of posttest.

Researchers also must consider how they will examine possible selection biases (Shadish et al., 2002). Selection biases are factors that may influence the group to which study participants are assigned. Since selection biases can bias experiments, researchers often depend on pretests to be certain that any selection biases are controlled. A pretest is any questionnaire or test that participants are required to complete prior to implementation of a treatment. While pretests may be an option in the design of true-experiments, pretests are essential in the design of quasi-experiments (Slavin, 2007). Pretests can be additionally utilized by researchers at the conclusion of a study to determine the effects of attrition on the study’s findings. Attrition occurs when a study loses participants during the course of an experiment for a variety of reasons (e.g., absenteeism, participants dropping out the study). It is important that researchers control for attrition, which is considered a threat to the internal validity of a study, as attrition may be a problem when the loss of participants impacts the results of a study. For instance, one small school district designed an experiment to determine the impact of a grade-to-grade promotion policy on test scores. At the conclusion of the study, findings were reported by many media sources detailing the massive test score increases in upper grades that resulted after the policy, which required that students pass a test to be promoted to the next grade level, was implemented. Unfortunately, only after the grade-to-grade promotion policy was implemented in several other districts did individuals begin to question the validity of the original study’s findings. In this situation, attrition biased the study results because the small school district had systematically removed many low-achieving students from the upper grades by retaining the students due to the grade-to-grade promotion policy. Thus, while the study concluded that the test scores of upper
grades seemed to massively increase, the increase may have been greatly influenced by the fact that many of the low-achieving students who may not have scored high on the tests had been retained in the lower grades (Slavin, 2007).

Another element of design that researchers must consider prior to beginning a quasi-experiment is the type of comparison group that will be utilized as a comparison to the group that will receive the experiment treatment. The purpose of a comparison group is so that researchers can determine what outcome may have occurred had the treatment group not received the treatment. It is important to note that a comparison group should be as alike as possible in as many dimensions as possible (Slavin, 2007). In the above example of a quasi-experiment of the effectiveness of the after-school program for increasing academic achievement, the principal selected a comparison group comprised of students who did not participate in the after-school program from schools demographically similar to the school that the after-school program’s participants attended. This will allow the principal to determine whether the increase of students’ academic achievement for students who were enrolled in the after-school program differ substantially when compared to students from similar schools who did not participate in the after-school program. Additionally, while this particular principal may have decided to use a nonequivalent comparison group there are several other types of comparison groups available for researchers to select. Some researchers select to use cohort controls to compare findings to a group that did not receive the experiment treatment. Cohort controls are often considered particularly useful since groups that move through a school often share similar backgrounds (e.g., a somewhat similar general socioeconomic status, the same age, etc.) (Shadish et al., 2002).

The final design element that researchers must consider when designing a quasi-experiment are specifics regarding the application of the study treatment (Shadish et al., 2002). A study’s treatment (the set of conditions applied to the treatment group studied by researchers to determine the effect of the treatment) (Slavin, 2007) can be manipulated to facilitate the process by which researchers determine whether a relationship exists between variables. One way in which researchers may select to manipulate the application of a treatment is by practicing a switching replication method. Switching replication occurs when, after an original study has concluded, the control group of the original study is given the treatment originally applied to the first treatment group. A better method practiced by researchers is to apply a treatment to multiple comparison groups at different times. Whereas reversed treatment requires that researchers apply a treatment at a different time that is expected to reverse the outcome achieved with the original treatment, removed treatment is a method in which researchers present a treatment then remove the treatment to study patterns of outcomes. The final way in which researchers often manipulate a treatment is by a method called repeated treatments. Repeated treatments occur when researchers introduce a treatment then remove the treatment only to reinstate the treatment afterwards as many times as feasible. This method is often called an ABAB design (A indicates the treatment while B indicates the removal of the treatment) (Shadish et al., 2002).

Types of Quasi-Experiments

One-Group Pretest-Posttest Design
Often new teaching strategies are toted as effective practices in studies that measured one group with a pretest, implemented a treatment manipulation, and then measured the same variable, as was measured with the pretest, with a posttest (Cohen, Manion, & Morrison, 2007). The following is an example of a one group pretest-posttest design.

A principal wonders whether a new reading program effectively increases the amount of words per a minute (WPM) second graders are able to read. The principal requires that teachers record the amount of WPM the second grade students can read at the beginning of the school year (this is an example of a pretest). Throughout the months of September and October teachers implement the new reading program in the second grade classrooms (this is an example of a treatment). Finally, to determine the effectiveness of the reading program at increasing second grade students’ WPM read, teachers retest the amount of WPM the second graders can read at the end of October (this is an example of a posttest). When results indicate that second graders’ WPM read increased substantially from August to October, can the principal determine that the new reading program effectively increased students’ WPM?

This design is often represented as: $O_1 \ X \ O_2$ with $O_1$ representing the pretest, $X$ representing the treatment implemented, and $O_2$ representing the posttest (Cohen et al., 2007). Could the principal in the example be justified in attributing second graders’ increase in WPM read to the new reading program that was implemented? At first many people may agree that it seems reasonable for the principal to make such an assumption, however, Slavin (2007) declares, “Pre-post test comparison are prone to so many errors and biases that they are rarely, if ever, justifiable” (p. 57). To understand Slavin’s statement, an individual must consider the numerous influences that might account for the experiment’s findings.

First, how can researchers be certain that the differences between pretests and posttests scores can be attributed to the treatment implemented? When a one group pretest-posttest design is utilized researchers cannot be certain that differences found are not from a variety of unaccounted for factors (such factors that exert influence on a study’s findings but are outside of researchers’ control are called extraneous variables) (Cohen et al., 2007). Another problem associated with interpreting findings from a one group pretest-posttest design is that researchers are unable, with such a design, to determine whether the differences noted on pretests and posttests are more or less than what should have been expected (Slavin, 2007). This problem is especially significant when one considers the nature of educational research. Since educators expect students to make academic gains throughout each school year, how can we be certain that gains found in particular experiments are greater than would have occurred had a particular treatment not been implemented? Using the appropriate scores helps determine if gains are made.

Another problem associated with interpreting findings from one group pretest-posttest experiments is that researchers, by simply requiring participants to complete a pretest, may impact posttest scores. This impact can be best illustrated when one considers the impact of requiring students to complete a spelling pretest several days before taking a final spelling test (the posttest). Educators often hope that by taking a spelling pretest on Monday that students will make note of words that they have spelled incorrectly and, thus spend a significant portion of time practicing the words that they spelled incorrectly on the pretest so that they learn to spell the words correctly and demonstrate their ability on the final test several days later. While educators
may hope that a spelling pretest positively impacts students’ scores on a spelling posttest, researchers do not want study participants’ completion of a pretest to impact their posttest scores in such a way. For example, researchers would not want a pretest on spelling words to impact students’ scores on the posttest if they were studying the effectiveness of a new spelling study routine. Rather, they would want to know if the new spelling study routine had an impact on the posttest scores.

One situation in which a one group pretest-posttest design may be useful is when pre-scores are available to educators that have remained stable for a long period of time. For instance, if a school were attempting to increase student attendance, educators might be justified in using a one group pretest-posttest design if the school has data that proves that the school’s attendance rate has remained fairly consistent for several consecutive years and that no other significant changes occurred during the specific year in which the school implemented a new program to increase student attendance (Slavin, 2007).

In the case of the earlier example in which a principal is attempting to interpret findings in which students’ word per minute (WPM) reading increased after the implementation of a new reading program, the principal must consider all of the factors that could possibly influence students’ reading. Perhaps another aspect of the classroom such as the materials provided during reading instruction or the teachers’ enthusiasm while teaching the new reading program in some way impacted the students’ reading scores. In this case, the principal cannot assume causality with a one group pretest-posttest experiment (Cohen, et al., 2007).

One-Group Posttest-Only Design

One of the least adequate of quasi-experimental designs is the posttest-only design that occurs when researchers implement a treatment without requiring study participants to complete a pretest to ascertain scores prior to treatment implementation while using a single group without a second group to use as a comparison group. For instance, consider the following example (Best & Kahn, 2006):

Mr. Lang, a high school history teacher, would like to know whether recent instructional changes he has implemented have increased students’ learning of the French Revolution. Mr. Lang administers an assessment to the class at the conclusion of the unit on which the majority of students score in the mid-90s. Can Mr. Lang conclude that the instructional changes were responsible for students’ high scores on the assessment?

Although Mr. Lang may be pleased that so many of the students scored well on the assessment, posttest-only experiments are difficult to interpret as they are subject to almost all threats of internal validity (Shadish, Cook, & Campbell, 2002). The fact that posttest-only designs lack not only a control group but also random assignment and a pretest, cause this methodology to be most often described as flawed (Cohen, Manion, & Morrison, 2007). For example, perhaps Mr. Lang’s students had been taught information regarding the French Revolution in a previous history class or the students may have learned about the French Revolution from the French foreign exchange student on campus when she discussed her ancestors’ experience during that time period in an article she’d written for the school’s newspaper. Mr. Lang cannot be certain of
the students’ level of understanding prior to his teaching since no pretest was conducted. Additionally, Mr. Lang cannot know whether his students’ learning increased due to the instructional changes he implemented nor whether the students would have scored similarly on the assessment had they been taught the material with the instructional practices Mr. Lang used prior to implementing instructional changes.

The Nonequivalent Control Group Design with Posttest Only

Non-equivalent control group posttest-only design involves an experiment in which a control group is utilized as well as a posttest. It is important to note; however, that control groups used by researchers in a non-equivalent control group posttest-only design are groups that have not been created through random assignment (Green, Camilli, & Elmore, 2006). Consider the following example:

Ms. Marshall is a K-12 music teacher who would like to determine whether a computer software program that she has recently implemented with gifted students advances their compositional skills. Ms. Marshall has decided to evaluate students’ compositional skills by scoring the level of complexity of student created compositions. Since Ms. Marshall had not conducted a pretest of students’ composition skills prior to implementing the software program (the experiment treatment), Ms. Marshall decides that she will use a non-equivalent control group posttest-only design as she has implemented the software program so far with only one section of her sixth grade gifted students. Thus, while one group of sixth grade gifted students will act as a control group for the experiment, the sixth grade gifted students that Ms. Marshall implemented the software program with will serve as the study’s treatment group. After students’ complete creating compositions, Ms. Marshall notes that sixth grade gifted students whose instruction had been supplemented with the software program created substantially more advanced compositions in comparison to the other group of sixth grade gifted students who were not taught with the program. Can Ms. Marshall determine that the computer software program advanced gifted students’ compositional skills?

Control groups, like the one used by Ms. Marshall in the example, provide information to researchers as to what would occur in the absence of the research treatment. For instance, Ms. Marshall can use the scores of students’ compositions who were not taught with the software program to determine the level of skill that students who had been taught with the program may have achieved had they not been taught with the program. Also, control groups are useful as they allow researchers to measure the size of the treatment effect. Of course, information provided by using a control group is greatly dependent on how similar the control group is in comparison to the treatment group (Green et al., 2006).

As with any situation in which a control group is used in a quasi-experiment design, researchers implementing a non-equivalent control group posttest-only design must consider selection bias (Green, et al., 2006). Selection bias becomes a concern anytime random selection is not used as the absence of random selection makes it difficult for researchers to determine whether treatment outcomes are due to a treatment or related to differences between groups. For instance, what if in the experiment Ms. Marshall conducted discussed earlier, Ms. Marshall had implemented the software program with a section of sixth grade gifted students that was an all male class, whereas
the control group was a class of all female sixth grade students? Can Ms. Marshall be certain that the results of her study are without bias? Groups can differ in numerous ways including in skill level and level of motivation. Since the ultimate goal of quasi-experiments that use a non-equivalent control group posttest-only design is to compare results between pretest and posttest scores of each group – the comparison group and the treatment group (Heiman, 1999), researchers must be certain that differences between the comparison group and treatment group used in the study do not unduly bias the results of a study (Shadish, Cook, & Campbell, 2002).

The Nonequivalent Control Group Design with Pretest and Posttest

The nonequivalent control group design with pretest and posttest has been described as “one of the most commonly used quasi-experimental designs in educational research” (p. 283, Cohen, Manion, & Morrison, 2007). This is often the case since students are naturally organized in groups as classes within schools and are considered to share similar characteristics (Best & Kahn, 2006).

The nonequivalent control group design with pretest and posttest is represented as:

Experimental Group: $NR \ 1 \ O \ X \ 2$

Control Group: $NR \ 1 \ O \ 2$

In this design $NR$ represents non-randomization, $1$ represents pretests, $X$ represents the treatment implemented, and $2$ represents posttests (Cohen et al., 2007). So while both the control and treatment group complete a pretest and posttest, the treatment group is the only group that receives the research treatment.

As with all other quasi-experiments, in this experimental design, groups are considered non-equivalent as groups are not randomized (Cohen et al., 2007). Nonequivalent groups specifically mean that participant characteristics may not be balanced equally among the control and experiment group. Also, non-equivalent groups mean that participants’ experiences during the study may differ (Heiman, 1999). More equivalent groups may be created through either matching or random treatment assignment. As matching is often impossible for practical reasons, researchers using nonequivalent groups should select samples from the same population, as well as selecting samples that are as similar as possible (Cohen et al., 2007). While nonequivalent groups may not be preferable, nonequivalent groups have been described as “better than nothing” (p. 320, Heiman, 1999).

Another advantage of the nonequivalent control group design with pretest and posttest is the pretest that both control and treatment groups complete. There are several benefits associated with pretesting including that the use of a joint pretest allows researchers to analyze differences that may initially exist between control and experiment groups which then allows researchers to adjust for such differences (Green, Camili, & Elmore, 2006). Another benefit of pretesting is that such tests tell about the magnitude of differences between control and treatment groups since researchers typically assume that differences between groups can be identified with pretests. This assumption requires that researchers look specifically at the size of the difference of pretest
scores. Smaller differences in pretest scores indicate that smaller differences may exist between control and treatment groups. Finally, pretesting also assist researchers while they statistically analyze data (Heiman, 1999).

What is the goal of nonequivalent control group design with pretest and posttest? How can researchers interpret findings of quasi-experiments using a nonequivalent control group design with pretest and posttest? Consider the following example:

A principal of an elementary school wants to determine whether a recently implemented school reform program significantly impacts student reading achievement as measured by the Tennessee Comprehensive Assessment Program: Achievement Test (TCAP: AT). As it is impractical to implement the school reform program with only some grades within the school, the principal decides to use a nearby school that serves students demographically similar to students at the principal’s school (these students will serve as the control group). At the end of the first year of implementation of the school reform program (the treatment), the principal analyzes students’ TCAP: AT scores before the school reform program was implemented (pretest scores) and the students’ TCAP: AT scores after the school reform program was implemented (posttest scores) and finds the following results:

<table>
<thead>
<tr>
<th></th>
<th>Control Group</th>
<th>Treatment Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean TCAP: AT Reading</td>
<td>40.02</td>
<td>39.67</td>
</tr>
<tr>
<td>Scores (pretest)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean TCAP: AT Reading</td>
<td>36.25</td>
<td>43.05</td>
</tr>
<tr>
<td>Scores (posttest)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Did the school reform program implemented in the example significantly impact students’ reading achievement as measured by the TCAP: AT? How should the principal go about interpreting the findings of his study?

Findings of studies that have used a nonequivalent control group design with pretest and posttest should be interpreted cautiously since control and treatment groups may differ due to selection bias (Best & Kahn, 2006). Researchers should not interpret study results by simply studying differences in posttest scores of the control and treatment groups since such differences could be attributed to differences in participants’ characteristics and/or differences in participants’ experiences during the experiment. Instead, results should be interpreted by comparing differences between each group’s pretest and posttest scores. To illustrate, let’s consider students’ mean TCAP: AT reading scores in the above example.
While the control group showed a decline in reading scores as measured by TCAP: AT from a mean of 40.02 to a mean of 36.25, the treatment group showed an increase in reading scores as measured by TCAP: AT from a mean of 39.67 to a mean of 43.05. Does this mean that students’ achievement scores were significantly impacted by the school reform program? To make such a determination researchers must complete a statistical analysis of the data with either an independent t-test or ANOVA analysis of variance (Heiman, 1999).
References


