Chapter 9

Two-Condition Experimental Studies

Up to this point the focus of the critiques has been mainly on descriptive studies. We looked at relationships between variables, predicting performance on the basis of several independent variables, describing a performance on the basis of related variables (factors), and differentiating performances on the basis of certain variables. Because no variable was actively manipulated, cause and effect conclusions could not be reached, nor were they sought. (There was one exception in Chapter 8: One study manipulated type of training to determine whether groups then would be differentiated.) The remainder of the evaluations focus on experiments: **Independent variables** are introduced and manipulated under the assumption that they will produce some change in behavior or make a difference. These will be the ultimate causes of behavior. The particular response or behavior being measured is the **dependent variable**. This will be the ultimate effect.

Researchers often have to decide whether to test different groups of individuals (between-groups design) or test the same group of individuals (within-groups design). If the focus of a study, the independent variable, is an individual characteristic (males vs. females, young consumers vs. older consumers), there is no choice; different groups of individuals have to be tested. If the focus is on relative distinctions people can make (5 vs. 10 seconds, hot vs. cold pain thresholds), there is no choice; the same group of individuals has to be tested. But in many other instances (effectiveness of different dosages of the same drug or of different

drugs, susceptibility to suggestion, effectiveness of speakers in attracting voters), there can be a choice, and the decision almost always is dictated by convenience (number of potential participants available, time to test a single individual, source of participants).

When different groups are tested, it is essential that we have some level of assurance that before the introduction of the independent variable, they are equivalent on whatever response will be measured. This can be accomplished in two ways. One way is to randomly assign the participants to the groups. The other is to first match the participants and then randomly assign them to the various groups. The most popular technique of **random assignment** is through the use of a table of random numbers. After the numerical basis for group assignment has been established (e.g., 1–10 = Group A, 11–20 = Group B), participants are assigned numbers, numbers are read from the table, and participants with corresponding numbers are assigned to the various groups (e.g., a participant whose number is 14 is assigned to Group B). The method assures that no bias is associated with group assignment. Randomization, however, does not guarantee that groups will not differ on an important variable that might just as easily account for results. It just makes it more likely that they will not differ. And the smaller the size of each group, the more likely that even randomization will not control for the extraneous variables that have to be controlled. Thus, even if pretest scores reveal no differences, it could be due to low statistical power rather than to equivalence.

An alternative approach to random assignment is **matching**. This can take several forms but always has the same goal: to ensure that groups are equivalent with respect to the dependent variable before treatment (the independent variable) is introduced. The ideal procedure is to give a test to all participants. Then, one can match participant for participant on the basis of identical scores on the pretest. This procedure, however, involves testing a host of individuals to get enough pairs of identical scores. Or we might devise two groups whose means and variances are alike to yield two matched groups. This is accomplished by ranking all scores, matching on the basis of rank orders, and then randomly assigning to the two groups. Of course, the matching task must be related to the response that will be measured in the study.

If the same group of participants is tested, it is crucial that the levels of the independent variables be introduced in such a way that the results are not due to practice or fatigue. This is accomplished by **counterbalancing** the introductions such that one half of the group receives A followed by B and the other half receives B followed by A; each level appears equally often at each stage of practice.

Two-condition designs are useful for answering one of two questions: Does a particular treatment work? Which of two treatments is relatively more effective? The first question can be answered by testing an experimental group that receives treatment and a control group that receives **placebo treatment**—that is, treatment that contains elements of the independent variable without its crucial characteristic (e.g., a sugar pill that is imbibed, an injection of saline, a sham operative procedure, a group rap session). If there is no carryover effect of treatment, the same group could be tested, but this seldom is the case. The second question can be answered by testing two groups or the same group, depending on the nature of the independent variable.

From a statistical point of view, the precise test used to evaluate effectiveness of the independent variable depends on the dependent variable. If scores are at least part of an interval scale (e.g., scores on a standardized test) and are not terribly skewed, a *t* test is most appropriate. This involves determining the probability that a means difference of the magnitude obtained would be obtained if treatment were not effective. If that probability is $\leq .05$, the obtained difference is attributed to the treatment. The *df* is determined by $n_1 + n_2 - 2$. In the event that a *t* test is inappropriate (e.g., when distributions are very skewed or variances are very different—both of which are likely to wash out any differences that might exist and lead to a Type II error), a nonparametric test can be used (e.g., Mann-Whitney U-test, median test).

When evaluating a two-condition study, there are two important questions to keep in mind. First, is the group equivalence that existed at the beginning of the study maintained? If testing occurs on more than one occasion, there is the possibility that attrition results in a selective loss—that is, remaining participants in each of the conditions may constitute two groups that are no longer equivalent. Second, aside from the independent variable, is there any other variable that can just as easily account for differences in performance between the two conditions? Thus, caution still is required regarding similarity in testing conditions as well as testing by a naive experimenter. If there were pretests and posttests, then other possible confounds include history, maturation, instrumentation, initial testing, regression toward the mean, diffusion of treatment, compensatory equalization, compensatory rivalry, resentful demoralization, and interaction effects. Finally, unless a study uses subhuman participants, there are ethical considerations that should be reported. (There are strict standards for the humane treatment of animals, as well.) All participants should have signed an informed consent slip, which also gives them the right to withdraw from the study at any time, without penalty. The list of caution factors, which is extensive, is found in Box 9.1.

We will evaluate two studies. I'll evaluate the first one with you, and you will do the second one.

Box 9.1 Caution Factors for Two-Condition Studies

- There was initial group equivalence by random assignment or matching.
- If groups were matched, the matching task was related to the dependent variable.
- If testing occurred on more than one occasion, there was no selective loss of participants.
- If a pretest–posttest design was used, nothing happened to one group that could affect performance (history).
- If a pretest–posttest design was used, one group did not change more than the other in a way that could affect performance (maturation).
- If a pretest–posttest design was used, the same instruments and testers were used.
- Participants were not selected because they were extreme scorers on a pretest (regression toward the mean).
- Performance on a posttest was not due to prior experience with a pretest (initial testing).
- Control participants could not learn about experimental treatment (diffusion of treatment; compensatory rivalry; resentful demoralization).
- Control treatment did not mimic experimental treatment (compensatory equalization).
- A threat to internal validity did not act selectively on one group (interaction effect).
- Tests used were reliable and valid.
- Testing was performed by someone who was naive with respect to purpose of the study.
- Testing was performed under uniform conditions.
- If a single group was used, conditions were counterbalanced.

STUDY EXAMPLE 9.1: "IMPACT OF AN INSTRUCTIONAL PROGRAM ON NURSES' ACCURACY IN CAPILLARY BLOOD GLUCOSE MONITORING"

This study is from the field of nursing and introduces instruction as the independent variable.

The Study

O'Neill, K. L., & Ross-Kerr, J. C. (1999). Impact of an instructional program on nurses' accuracy in capillary blood glucose monitoring. *Clinical Nursing Research, 8*(2), 166–178. Copyright © 1999 by Sage.

The increasing level of responsibility assumed by nurses in a variety of practice settings inevitably raises questions of accuracy of performance of techniques. The sheer volume of procedures that nurses are expected to perform competently has risen dramatically over the past two decades, as has the challenge of mastering the full range of nursing techniques for a particular area of practice. . . . Nursing managers responsible for quality assurance must be attuned to the skills required of nurses, because new nursing skills emerge regularly from advances in treatment and care and because new skills are continually being transferred to nursing from other disciplines. Of concern are such issues as how to ensure that new and experienced nurses are competent in the performance of particular techniques and how to assist them to maintain their competence on an ongoing basis. . . .

The importance of practitioner competence in performing nursing procedures has long been recognized in schools of nursing and in staff development, and nursing educators have sought to assist nursing students and registered nurses to gain competence in performing an array of procedures. In nursing departments of hospitals and home care agencies, a process of credentialing using certification for designated procedures has been favored to ensure that nurses have received proper instruction and practice in learning to perform particular techniques. . . .

The inspiration for this research project emerged within the context of quality of practice and professional accountability in clinical nursing practice. The intent was to investigate the impact of an instructional program on the accuracy of performing a common nursing procedure in health agencies, that of visual readings of glucose oxidase strips to monitor capillary blood glucose [BG] levels [of patients with diabetes]. Even though the strips may be read accurately by passing them through a reflection meter, the expense of supplying these meters and maintaining their calibration for all areas of an agency where they are used may restrict their use to specialized or intensive care units. As a result, nurses in most agencies are expected to be able to skillfully perform visual readings of glucose oxidase strips.

Background of the Study

In an average 500-bed hospital, nurses can be expected to perform 400 to 500 glucose oxidase . . . procedures per week, the results of which form the basis of critical clinical decisions, including insulin dosage. . . . Although products vary, the procedure is based on the principle of a color change in a reagent strip containing glucose oxidase coupled to a chromosome, in response to the glucose in the capillary blood sample. . . .

Ease of carrying out the procedure, speed of results, and reliability established by comparing strip readings to laboratory findings have led to wide use of this method of capillary glucose monitoring by nurses at the bedside. . . . Studies of measuring capillary [BG] using glucose oxidase strips have shown remarkable levels of accuracy in the presence of adequate education and frequent performance. . . . Specialized nurses and laboratory workers have shown consistently high correlation coefficients of .92 to .98 compared to reference laboratory values. . . .

The reliability of glucose monitoring by bedside nurses, when measured on a periodic basis, has been less than optimal. Two studies found accurate readings only 50% to 60% of the time when compared with concurrent laboratory findings. . . . Instructional programs have been shown to be associated with increased accuracy, and several studies have documented acceptable levels of accuracy as a result of ongoing educational programs. . . . Despite the frequent use of glucose oxidase strips, there are, however, continuing concerns about the accuracy of monitoring. Although prior to the development of this technology, . . . [BG] monitory has been a function traditionally performed by laboratory staff. In many cases, nursing departments have assumed this function without the requisite quality control program that has been the hallmark of operating procedures in a laboratory setting.

Research Question

The primary research question in this study was to determine the impact of an instructional program on visual readings of glucose oxidase strips to monitor capillary glucose levels on the accuracy of such measurements by nurses employed on general medical units in a medium-size, urban, community general hospital. Other questions explored accuracy of performing glucose oxidase strip readings by nurses on the four medical units in the study hospital and the effect of the time lapse between the completion of the instructional program and the performance of the procedures on accuracy of performance of the technique.

1. What was the rationale for the study?

Nurses today have more responsibilities and require more skills, some of which have come from other disciplines. There is a need for assurance that new and experienced nurses are competent in performing these skills. Some nursing departments in hospitals and home care agencies require that nurses be certified to demonstrate their proficiency.

This study focused on quality of practice of a particular skill, visual reading of glucose oxidase strips to obtain a measure of capillary [BG]. Meters are available but are limited in availability and require maintenance. Therefore, most readings are visual.

In a 500-bed hospital, nurses typically read 400 to 500 strips per week (for patients with diabetes). Accuracy is important because it determines insulin dosage. Specialized nurses are very accurate; bedside nurses are less accurate. Specialized training is beneficial, but the technique is not always rigorously taught.

2. What was the purpose of the study?

The primary purpose was to evaluate the effectiveness of an instructional program on visual reading of strips by nurses in general medical units of a hospital. Secondarily, accuracy of nurses on four medical units was assessed as well as the effect of time since training on accuracy.

Method

Design

This was an experimental study in which the unit of analysis was glucose oxidase strip readings performed by nurses who consented to participate and were randomly assigned to the experimental or control groups. The instructional program served as the independent variable and the accuracy of performance as the dependent variable. In exploring the time lapse between instructional input and practice, the time period served as the independent variable and accuracy of performance the dependent variable.

Sample

Eighty-four nurses employed on four general medical units of the study hospital initially consented to participate in the study and were randomly assigned to experimental and control groups. All patients hospitalized on these units who had an SMA-6 or laboratory [BG] test ordered by their physician during the data collection period were asked to participate in the study and written consents were obtained from 110 patients, some of whom participated more than once.

▶ *(Note that nurses consented [as did patients] and that the groups were formed by random assignment.)*

Procedure

Each participating nurse was asked to complete two glucose oxidase procedures on patients with laboratory [BG] or SMA-6 ordered and to notify the laboratory staff in advance that a concurrent glucose oxidase strip reading was to be done. The strip reading was carried out within 1 minute of the time the blood was drawn by the laboratory staff. Data forms with . . . nurse's code number, date, time, and strip reading result were completed by the nurses after each test and placed in a designated box on the unit. . . . The concurrent laboratory [BG] values were marked on the data forms as soon as the results were obtained. Procedures

were the same for nurses in the experimental and control groups, except that the nurses in the control group did not receive the educational program. At the conclusion of the 3-month data collection process, 36 nurses from the experimental group had performed a total of 70 glucose oxidase strip readings, whereas 34 nurses from the control group had performed 68 similar procedures. Six data sheets on procedures had to be discarded because of either incorrect timing or missing information, and four nurses withdrew from the study due to resignation or transfer.

▶ *(Note that the final N is 70 and that only 10 more of the original 84 are accounted for. Also, the instruction program is not described, nor is its length.)*

3. What was the general procedure?

Nurses were randomly assigned to experimental or control groups. When a blood test was required on a patient (who consented to take part in the study), two procedures were conducted. One strip was read by the nurse and the other by a trained technician. Each nurse filled out required forms for each reading and was given a code number. All completed forms were kept in a designated box. Experimental participants received instruction; control participants did not.

4. What was the final sample?

After a 3-month data collection period, 70 of the original 84 nurses completed the study. There were 36 in the experimental group and 34 in the control group; six data sheets had to be discarded, and four nurses withdrew. This leaves four unaccounted for.

5. What information is lacking?

Nothing is presented about the training procedure, number of nurses at each session, length per session, number of meetings, times of meetings, nor what control participants did at the time.

Reliability and Validity

Randomized assignment of nurses to experimental and control groups minimized the effects of history and maturation, and such external factors as educational background, work experience, and exposure to previous educational

orientation programs were assumed to be distributed equally across the groups. Thus, differences between groups could be attributed to the impact of the intervention.

▶ *(Note that this is not unequivocal because experimental participants may have received more attention during training. We don't know what control participants did. Moreover, although nothing more was said about the loss of participants, it was such a small number that selective loss is unlikely to be a factor.)*

Although four participants withdrew from the study, experimental mortality was not a significant factor. . . .

The accuracy and precision of the laboratory glucose test value was assured because of the well-established quality control program in the laboratory. . . .

Data Analysis

Data were analyzed using . . . *t* tests. . . . Results for the experimental and control groups were then compared on the basis of a one-tailed *t* test for independent samples. In analyzing the effect of the time interval between the instructional program and the performance of the test, an arbitrary period of 3 weeks was established. Results for the experimental group were subdivided to those readings that were done within the 3-week period beginning at the time of the instructional program and those performed outside the 3-week time frame. . . . Differences between the groups [were] compared on the basis of a one-tailed *t* test. Significance level for all statistical tests was established at $\alpha = .01$.

[Presumably, the instruction period lasted 3 weeks. Moreover, half the experimental readings were done during the time of instruction, and half were done after the 3-week period. But we don't know how much later, because data were collected over a 3-month period. Finally, because this was an evaluation study, t tests should have been two-tailed. The only saving grace is that alpha was .01, much more stringent than .05.]

6. Readings by experimental participants were divided into two groups: 35 readings made during the 3-week instructional period by one half the group and 35 made after the instructional period by the other half. What difficulty might this produce for the analysis of results?

Time since training (recall period) had to be variable for the second group of nurses because the entire data collection period was 3 months. Increased variability

could decrease the possibility of finding a significant difference, which in this case would be a favorable finding.

Results and Discussion

Effect of Educational Program on Reading Glucose Oxidase Strips on Accuracy of Performance

The primary research question in this investigation was whether an instructional program offered to nurses would result in more accurate performance of the technique or higher correspondence between visual glucose oxidase strip readings and laboratory [BG] values. Results indicated that the percentage difference between the glucose oxidase strip and laboratory values in the experimental group was significantly lower than the same percentage difference in the control group. The *t* test used to compare the mean percentage differences for averaged Test 1 and Test 2 data for each of the experimental and control groups yielded $t(68) = 5.16$, a figure that was far above the critical value of 3.35 for a one-tailed probability of .01 (see Table 9.1). This confirmed that the experimental group was significantly more accurate than the control group in performing glucose oxidase strip readings using concurrent laboratory values as the standard. Because the groups were randomly assigned and because the only known difference between the groups was their exposure to the intervention, the difference between the groups was attributed to this intervention.

Table 9.1 Differences Between Blood Glucose Monitoring Recordings Obtained by Nurses and Corresponding Laboratory Values for Experimental Control Groups

Variable	*Number of Cases*		*Separate Mean*	t *Value*	*Degrees of Freedom*
PD1	Group 1	36	11.44	3.64[a]	68
	Group 2	34	71.53		
PD2	Group 1	34	8.11	3.68[a]	66
	Group 2	34	41.74		
$PD_{\bar{x}}$ (mean of PD1 & PD2)	Group 1	35	9.87	5.16[a]	67
	Group 2	34	56.63		

Note: PD = Percentage difference between chemstrip recording and laboratory BG recordings. $PD_{\bar{x}}$ = mean of PD1 and PD2 for each of Groups 1 and 2.

a. $p < .01$ for a one-tailed test.

▶ *(Note that the results are significant with a two-tailed test at* $p < .02$*.)*

7. Look at Table 9.1. Is there any evidence of a practice effect for the control nurses?

Yes. On their first test, they averaged a 71.53% difference in accuracy between their readings and laboratory readings of the strips. By the second test, the average difference was reduced to 41.74%, a difference of 29.79% compared with a difference of 3.33% by the experimental nurses. Put another way, by the second test, the difference in accuracy between both groups was almost half of what it was for the first test.

8. Table 9.1 also shows that the values of *t* were virtually the same for both tests. What does this suggest about variability?

Because the difference in means was smaller for the second test (33.63 vs. 60.09), *t* could only be the same if variability were less. For the first test, 60.09/3.64 = 16.508; for the second test, 33.63/3.68 = 9.139.

Effect of the Time Interval Following Completion of the Educational Program on Accuracy of Performance of Glucose Oxidase Strip Readings

The experimental group was further subdivided into two groups by the time lapse between the completion of the instructional program and the date of performing the capillary [BG] strip reading. One subgroup included those who performed strip readings within 3 weeks of the instructional program, whereas the other included those recordings that were performed at a greater than 3-week time interval. The total number of recordings in the experimental group was 70, 35 of which were done within the 3-week time frame and 35 outside this interval. The groups were then compared to determine differences in accuracy of performance.

▶ *(Note that the data were divided into subgroups. No mention is made about the nurses other than that one group made the first 35 readings and the other group made the second 35 readings. If, by chance, the second group had better retentive ability than the first group, their accuracy could well match that of the group tested during training. If the subdivision had been made initially, with one group randomly assigned to test first and the second group to test later, this and other individual characteristics would be ruled out.)*

. . . . The one-tailed *t* test did not yield a significant difference between groups (see Table 9.2). This finding indicated that the length of the time interval between completion of the instructional program and performance of the test did not have a significant effect on accuracy of performance of the strip reading. Thus, the nurses retained considerable accuracy of their performance on the test when they did not have an opportunity to practice the technique for more than 3 and less than 6 weeks following completion of the instructional program. . . .

Table 9.2 Differences Between Blood Glucose Monitoring Recordings Obtained by Nurses and Corresponding Laboratory Values According to Time Lapse Between Instructional Program and Test Performance

Variable	*Number of Cases*		*Separate Mean*	*r Value*	*Degrees of Freedom*
Chemstrip value	Group 1	35	7.08	2.27	34
	Group 2	35	8.90		
Lab value	Group 1	35	7.19	1.98	34
	Group 2	35	9.07		
Difference between chemstrip and lab values	0–3 weeks	35	0.67	0.94	34
	3+ weeks	35	0.97		

Note: None of the differences are significant at the .01 level for a one-tailed test.

▶ *(Note that this time frame includes 9 weeks. if data were collected for 3 months, at least some of the control nurses were tested during this additional 3-week period.)*

9. What do the results in Table 9.2 indicate?

Visual readings were just as accurate within the 3-week training period as they were at least 3 weeks after training.

10. Can any factor other than no loss in retention account for the lack of differences?

Yes. Under the assumption that the two groups were not formed by random assignment, the second group may have been superior in ability to recall. The point is, if this is possible and roles of the nurses were reversed, the "second" group might have shown a loss in retention.

Accuracy of Glucose Oxidase Readings

Another purpose of the study was to determine the accuracy of glucose oxidase readings being done by nurses staffing the medical units of the study hospital. Percentage differences were calculated between the nurses' first and second glucose oxidase strip readings and the corresponding laboratory [BG] values. For the first set of readings, the percentage difference values ranged from 1.83 to 363.33 ($\bar{x}$= 71.53) . . . n = 34. . . . Similar results were found for the second test completed by each of the 34 nurses in the control group. Here, percentage differences between the nurses' strip reading and the laboratory [BG] value ranged from 1.14 to 236.36 ($\bar{x}$= 41.75). . . . Results indicate that the glucose oxidase strip readings by the nurses were highly unreliable as measured against the standard of the laboratory [BG] value.

▶ *(Note that nothing is stated about the basis for selecting the 34 sets of readings. Moreover, the means were virtually identical to those of the control group.)*

Implications of Findings for Practice

Analysis of the control group data confirmed that the initial concerns about accuracy of nurses' performance of the glucose oxidase strip readings without adequate instructional support had been valid. . . .

Although retention of knowledge and skills learned in the educational program was maintained beyond a 3-week time interval in this study, . . . careful examination of study results revealed that there was some decline in accuracy of performance of the procedure in the subgroup when there was a longer time interval between completion of the program and performance of the test. Even though this difference was not a significant one, it does not seem particularly surprising that performance appeared to be enhanced by the presence of a relatively short time interval between theory and practice. . . .

Finally, the powerful impact of a well-designed and implemented instructional program on improving the accuracy of performance by nurses of visual capillary glucose strip readings has been demonstrated in this study.

11. What were the major conclusions reached by the authors?

First, the instructional program had a major impact in improving accuracy in reading glucose oxidase strips. Second, the effects of training were maintained for as long as 3 weeks after the program ended. Third, nurses in medical units cannot read the strips with any reasonable degree of accuracy.

12. Are the conclusions justified?

Despite the lack of crucial information regarding the nature of the training program and treatment of the control group (e.g., sham meetings without training would have been laudable), the first conclusion probably is tenable. Reading strips requires skills that could not be "picked up" even if a Hawthorne effect were operating. Nonetheless, the improvement shown in the control group by the second reading shows that aspects of the skill may be learned without formal training (and, presumably, without feedback).

The second conclusion may or may not be justified. If both experimental groups were equal with respect to their ability to profit from training and to retain the skills, then the conclusion is tenable. There is no basis, however, for assuming this equality, because we don't know why certain nurses were tested first.

Finally, the third conclusion would be justified if the 34 nurses' readings were randomly selected. Given the large ranges in accuracy, the sample does not appear to be biased. The extreme low ends of the ranges, however, also suggest that some nurses were skilled at reading strips.

13. To which population do these results generalize?

The results generalize to nurses in general hospitals in Alberta, Canada (where the study took place), who have experiences comparable to those of the current sample of nurses.

STUDY EXAMPLE 9.2: "THE EFFECTS OF A SHORT-TERM EXERCISE PROGRAM ON MOVEMENT, PAIN, AND MOOD IN THE ELDERLY"

This study, also from the field of nursing, is concerned with the effects of a special kind of exercise on various aspects of well-being of elderly individuals.

The Study

Ross, M. C., Bohannon, A. S., Dabis, D. C., & Gurchiek, L. (1999). The effects of a short-term exercise program on movement, pain, and mood in the elderly. *Journal of Holistic Nursing*, *17*(2), 139–147. Copyright © 1999 by Sage.

As the percentage of the U.S. population grows older, holistic interventions that prevent falls and their debilitating injuries are increasingly important.

Many studies verify the age-related changes in the body that impair balance and strength as one grows older and thus place the elderly at an increased risk of falling. . . . Reduction of visual, ankle, and foot inputs results in a loss of balance for up to 50% of the elderly, in contrast to fewer than 10% of younger participants. . . . The elderly also experience a decline in muscle mass and strength. Therefore, it is imperative to identify interventions to lessen these physical changes in the elderly to prevent incidences of falls by them. Tai Chi is one such intervention that has been identified as a safe and appropriate exercise for the elderly.

Tai Chi is not a vigorous type of exercise and can be an excellent choice for the older adult who lacks physical conditioning or the self-confidence required by more traditional exercise programs. . . . It . . . has been used for centuries in China as an exercise among elderly citizens. It is practiced for agility, balance, posture control, and mind–body interactions. . . . Tai Chi consists of a series of slow, purposeful movements that involve turning, shifting one's weight from one leg to the other, bending and unbending the legs, and [doing] various arm movements. . . . Tai Chi also has been described as a combination of deep diaphragmatic breathing and relaxation with slow and gentle movements, isometric and isotonic, and maintenance of good posture. . . .

The positive effects of various forms of exercise, especially activities that include body scan techniques or focused breathing, have been thoroughly documented in mind–body studies. Mindfulness activities such as these have positive effects on mental state and physical health. . . .

The physical and psychological therapeutic effects of Tai Chi for the elderly have been the subject of several recent reports. . . . Tai Chi exercise may benefit the elderly, because the exercise incorporates elements of strengthening, balance, postural alignment, and concentration. All four of these factors are essential to prevent fall[s]. . . . Tai Chi focuses on postural alignment, muscle strengthening, and improves the elderly person's perception of body position in space.

Furthermore, inasmuch as the majority of falls occur when the elderly person is walking, turning, or climbing stairs, balance is essential. Tai Chi includes single-stance balance training, which improves coordination and may decrease the elderly individual's risk of falls. . . . For osteoarthritis and rheumatoid arthritis patients, Tai Chi has been shown to have significant effects on lessening pain and tenderness and improving joint flexibility. . . .

. . . Improvements in self-esteem, self-confidence, sleep, and depression have been seen as a result of participation in Tai Chi. . . . In studies . . . Tai Chi students were less anxious and depressed following only 12 weeks of training.

. . . As a safe and enjoyable program, Tai Chi may have timeless implications for the health of today's elderly.

The purpose of this research was to conduct a pilot study to evaluate the effect of a short-term Tai Chi program on flexibility, balance, sway, pain, and mood in the elderly.

Research Questions

The research questions for the study were

1. Is there a difference in the degree of flexibility in the extremities of elderly persons after participating in a Tai Chi program?
2. Is there a difference in balance and sway in elderly persons after participating in a Tai Chi program?
3. Is there a difference in the amount of pain experienced, with active and passive range of motion of specific extremities of elderly persons, after participating in a Tai Chi program?
4. Is there a difference in the mood of elderly persons after participating in a Tai Chi program?

Method

Sample Characteristics

A convenience sample of 17 volunteers (2 males, 15 females) [aged 68 to 92 years], who were living independently in a public housing facility, participated in the study. Participants were solicited through written flyers and personal invitation delivered to their residences. All participants were screened to determine their physical acceptability for the study. Chronic illnesses, blood pressure, pulse, weight, respirations, medications, previous occupations, and current amount of routine exercise or activity were recorded, as a part of this screening. Those with any conditions, medications, or occupations that may have confounded the findings or may have presented a risk to the participant were eliminated. The participants were ambulatory and English-speaking. The sample included one Black and one Hispanic female. None of the participants reported current use of pain-relieving medication other than NSAIDS or acetaminophen. Those currently in occupational or physical therapy were excluded. Four participants were not included in the final results, because they did not attend at least 90% of the Tai Chi classes.

Procedure

Informed consent was provided verbally and in written form, according to the [Institutional Review Board] (IRB) protocol. The participants were interviewed and pretest baseline measures were taken in the following sequence:

1. A brief pencil-and-paper multifactor instrument, the Multiple Affect Adjective Check List Revised (MAACL-R), was completed by each participant, as a measure of mood. The MAACL-R consists of 132 adjectives describing feelings.... The MAACL-R provides valid and reliable data on five affects: anxiety, hostility, depression, positive affect, and sensation seeking. The MAACL-R is written at no higher than the 8th grade reading level.... The participants were asked to simply put an X by the words that described how they generally felt (trait form). The MAACL-R took approximately 5 minutes for the participants to complete.... The MAACL-R was hand-scored to provide a rating on the five scales. Numerous studies have established both validity and reliability of the MAACL-R with adults....

2. A "sit and reach" test of flexibility was performed. The best effort of three attempts was recorded, in centimeters.

3. A single-stance balance test was accomplished. In this test, participants were instructed to lift one foot just off the floor, focus on a point on a nearby wall in front of them, and stand that way for 60 seconds. The test was terminated when the foot touched the ground, rested against the other leg, or the participants raised their arms for balance. Pretest marching in place was used to produce a more natural foot placement, in terms of foot abduction and stance width. Research assistants stood on each side of the participants for safety. The single stance was timed by the use of a standard [handheld] stopwatch.

4. Sway was measured by the tandem walk (heel-to-toe walk). Participants were instructed to walk along a wide line for 15 steps without swaying off the line or grasping an attendant for balance. Each participant was given two trials. The score of his or her best performance on each test was then recorded. If the participant reached the goal of 60 seconds or 15 steps on the first trial, the second trial was not attempted.

5. Active range of motion of upper and lower extremities was performed by each participant, as an additional test of flexibility. A certified athletic trainer used a standard goniometer to measure range of motion, in degrees of flexion and extension of the elbow, shoulder, knee, and hip. At the end of each range of motion activity, the participant was asked to rate the degree of discomfort, if any, on visual analog (0 to 10) scale.

Identical measures, techniques, and data collections were used at the end of the 8 weeks of the Tai Chi sessions. . . .

Each Tai Chi class was conducted by a certified and highly experienced Tai Chi instructor. The classes took place three times each week and included 10 minutes of preparation and wrap-up administrative time and 50 minutes of actual Tai Chi movements. Registered nurses were in attendance at all classes to observe participants for any distress.

Data Analysis

Pretest–posttest comparisons were made on all measures for the group and individual participants. Demographic data were examined for spurious findings; however, none were observed. Statistical comparisons, including frequencies and *t* tests, were used to determine statistical differences. . . .

Results

Because this was a pilot study with a very small sample, the results must be interpreted conservatively. Statistically significant improvements (*t* tests), at the .05 level, were found in posttest measures of pain (visual analog scale) and mood (MAACL-R). Mean pretest perceived pain was 1.5473 ($SD = 2.279$), and that for posttest was 0.6945 ($SD = 1.165$). The difference was significant, $t(10) = 2.35$, $p = .041$.

Composite scores on the MAACL-R demonstrated improvement, although individual subscale scores did not show a significant change. The mean pretest measure of mood was 1.0909 ($SD = 1.300$), and the mean posttest measure was 0.6364 ($SD = 1.206$). The difference was significant, $t(10) = 2.19$, $p = .053$.

Although not statistically significant, all participants had measurable improvements in flexibility (goniometer measurements), balance (single-leg stance), and sway (length of time walking a line). However, when pre- and posttest scores on each individual were compared, six participants had significant improvement ($p = .05$) in balance, sway, and flexibility.

By nature, a pilot study is used to test research methods that may be improved in a comprehensive study. The research team quickly found that the elderly participants needed frequent reminders and encouragement to attend the beginning classes, until the group developed some cohesiveness and skills in the Tai Chi movements. . . . Testing appointments for elderly participants had to allow time for slow movement between testing areas, privacy for each test, and comfortable seating. Testing appointments prior to 10:00 [a.m.] were not selected by any of the participants, and food snacks after testing were well received.

Conclusions and Recommendations

This pilot study supports the use of Tai Chi as a safe, viable method of exercise for the elderly that has some statistically significant health results. As such, Tai Chi is a good choice for senior activity centers and senior residential programs.

In this pilot study, positive changes in mood, pain perception, flexibility, balance, and sway were demonstrated with even a limited program of Tai Chi exercise in older adults. . . . Expanded studies should include larger samples of men and women and racial and ethnic diversity. Also recommended are studies with an increase in the number of weeks of the Tai Chi classes. . . .

Tai Chi may be a safe, viable method of exercise for the elderly that has significant positive health results in four dimensions of human life—physical, emotional, mental, and spiritual. . . . Long-term studies of the effect of Tai Chi on the incidence of falls, depression, and analgesic use also may be important.

CRITIQUE OF STUDY EXAMPLE 9.2

1. What was the rationale for the study?
2. What was the purpose of the study?
3. What was the nature of the sample?
4. What demographic information, not given directly, can be surmised about the participants?
5. What was the general procedure?
6. What were some good features of the design?
7. What were some questionable features of the design?
8. How many sets of scores were included in the final analysis of results?
9. What were the findings regarding posttest pain and mood?
10. What were the findings regarding flexibility, sway, and balance?
11. What other reactions to the training sessions were noted?
12. What did the authors conclude on the basis of this pilot study?

13. What other variables could account for the results?
14. What group and design are essential to rule out these confounds?

For answers to these questions, see page 368.

SUGGESTED READINGS

Ary, D., Jacobs, L. C., Sorensen, C. K., & Razavieh, A. (2009). *Introduction to research in education* (8th ed.). Orlando, FL: Harcourt Brace.

Best, J. W., & Kahn, J. V. (2005). *Research in education* (10th ed.). Boston: Allyn & Bacon.

Harris, M. B. (1997). *Basic statistics for behavioral science research* (2nd ed.). Boston: Allyn & Bacon.

Isaac, S., & Michael, W. B. (1995). *Handbook in research and evaluation: A collection of principles, methods, and strategies useful in the planning, design, and evaluation of studies in education and the behavioral sciences* (3rd ed.). San Diego, CA: EdITS.

Lehman, R. S. (1991). *Statistics and research design in the behavioral sciences*. Belmont, CA: Wadsworth.

McMillan, J. H., & Schumacher, S. (1997). *Research in education: A conceptual introduction* (4th ed.). White Plains, NY: Longman.

Chapter 10

Single Classification Studies

The two-groups or two-conditions situations are limited. They inform you about whether or not a qualitative independent variable works or which of two effective treatments is more effective. If more than one level of the independent variable is introduced, then additional questions can be answered about relative effectiveness of each level. Because the treatment is classified in only one way, we call these **single classification studies**. At a minimum, three levels can be studied. They can be two different therapies versus a control condition or two different drugs versus a control, or even three different therapies or drugs. (If effectiveness of each has been established and relative effectiveness is being investigated, a control condition may not be necessary.) If the summary data are means, the type of analysis depends on whether separate groups have been tested at each level or the same group was tested at each level. Keep in mind that we are dealing with qualitative (distinct, noncontinuous) variables. When independent variables are quantitative (continuous, such as levels of anxiety), data are analyzed by multiple regression.

When separate groups are tested, participants are randomly assigned to the different levels of the independent variable. By this procedure, we have some assurance that, before treatment, participants at each level initially are equivalent on the dependent variable and other characteristics related to it. Therefore, differences in performance after treatment can be attributed to the independent variable. If no mention is made about random assignment, nonequivalence among groups looms as a possible confound. (Another procedure, used with small *n*s, is to match the participants on the dependent variable, or on one related to it, and then assign the conditions to the matched groups.) Likewise, if testing occurs on more than one occasion, you need assurance