Counterbalancing

In: The SAGE Encyclopedia of Communication Research Methods

By: Elena F. Corriero
Edited by: Mike Allen
Book Title: The SAGE Encyclopedia of Communication Research Methods
Chapter Title: "Counterbalancing"
Pub. Date: 2018
Access Date: August 12, 2020
Publishing Company: SAGE Publications, Inc
City: Thousand Oaks
Print ISBN: 9781483381435
Online ISBN: 9781483381411
DOI: https://dx.doi.org/10.4135/9781483381411
Print pages: 278-281

Counterbalancing is a procedure that allows a researcher to control the effects of nuisance variables in designs where the same participants are repeatedly subjected to conditions, treatments, or stimuli (e.g., within-subjects or repeated-measures designs). Counterbalancing refers to the systematic variation of the order of conditions in a study, which enhances the study's interval validity. In the context of experimental designs, the most common nuisance factors (confounds) to be counterbalanced are procedural variables (i.e., temporal or spatial position) that can create order and sequence effects. In quasi-experimental designs, blocking variables (e.g., age, gender) can also be counterbalanced to control their effects on the dependent variable of interest, thus compensating for the lack of random assignment and the potential confounds due to systematic selection bias. Counterbalancing does not eliminate order or sequence effects, but it distributes them evenly across all experimental conditions so that their influence is "balanced" and does not confound the main effects due to the independent variables.

This entry first discusses the importance of counterbalancing in relation to order and sequence effects. Second, different counterbalancing designs are explained, addressing the distinction between complete and incomplete counterbalancing, and providing examples of the major incomplete counterbalancing techniques. Finally, the application of Latin squares design to counterbalancing is considered.

Order and Sequence Effects

The goal of counterbalancing is to ensure internal validity by controlling the potential confounds created by sequence and order effects. A sequence effect (e.g., practice) occurs when responses to a condition are influenced by the sequence in which conditions are presented. Order effects occur when the position a condition occupies in the research protocol (e.g., 1st, 2nd) influences the response (e.g., fatigue effects). Suppose that participants in a laboratory experiment are asked to interact with a remote partner three times, each time through a different channel (video, chat, and text message), and measured after each interaction on feelings of self-efficacy. The researcher's goal is to assess which channel obtains the highest ratings of perceived self-efficacy. Further, suppose that all participants interact through the three channels in the same order: first video, then chat, and finally text message. Because the sequence is the same for all participants, low self-efficacy attributed to video cannot be univocally attributed to the channel, since scores may have been tainted by the participants' lack of practice with the study protocol (order effect); similarly, low ratings for text message (the last condition) may reflect fatigue due to a long experimental session (order effect), or may be due to a comparison with the preceding condition (sequence effect). Thus, the main effects of channel (the independent variable) will be confounded by order and sequence effects unless counterbalancing is used.

Counterbalanced Designs

Counterbalancing can be obtained through different designs. The major distinctions are between intrasubjects and intersubjects designs, and complete and incomplete designs. The first distinction refers to exposure of participants to the conditions. Intrasubjects counterbalancing allows for order and sequence effects to be balanced within subjects by exposing each participant to all conditions multiple times and in different orders, and is obtained through either ABBA counterbalancing or block randomization.

ABBA counterbalancing requires that each subject is exposed to conditions in some random order (AB) and subsequently exposed to the reverse order (BA). It can be employed when practice effects are assumed to be linear, and when it is not expected that subjects may form expectations leading them to change their behaviors (anticipation effects). If practice effects may be nonlinear, or anticipation could occur, block randomization is preferred. In this case, multiple blocks are created, with each block containing all the conditions in a different randomized order, and the participant is repeatedly measured on different blocks. Block randomization may be impractical if there are many conditions, because the subject must be tested on multiple blocks.

Conversely, in intersubjects counterbalancing, each participant is exposed to each condition only once, with order and sequence effects thus balanced across subjects. Intersubjects and intrasubjects designs can also be combined.

The distinction between complete and incomplete designs refers to whether all the possible permutations of conditions or treatments are used (complete), or only a subset of all the possible permutations is chosen (incomplete). Choice of design will depend on theoretical considerations about potential confounds, on the researcher's focus, and on empirical considerations including the number of conditions and of participants available. Counterbalanced designs allow the researcher to isolate the main effects due to condition and control for order and sequence effects only if there is no interaction between the procedural variables (time, position) and the independent variables.

Counterbalancing is based on the assumption that order and sequence effects are the same regardless of the specific sequence involved and that there is no asymmetrical or differential transfer of order or sequence effects (carry-over effect). This would occur if practice or fatigue effects for the sequence AB were different from the sequence BA. In this case, counterbalancing would not remove the confounds due to order or sequence. Although this assumption can be tested through analysis of variance, if the researcher expects carry-over effects, a between-subject design is recommended over a within-subjects one.

Complete counterbalancing is considered the best option for within-subjects designs, and should be always used whenever enough participants are available. It is obtained by employing all possible condition permutations and assigning an equal number of participants to each of the sequences. For instance, a three conditions (k) experiment (A, B, C) would produce 3! or $3 \times 2 \times 1 = 6$ different combinations: ABC, ACB, BAC, BCA, CAB, CBA. Table 1 shows the application of complete counterbalancing to the previous example of a computer-mediated communication study. Group 1 is measured on self-efficacy three times: after a video interaction, after a chat interaction, and after a text-message interaction. For each of the other groups, the sequence is rearranged so as to use all of the possible permutations of the three channels. Obviously, the number of participants should be a multiple of the number of sequences tested.

However, complete counterbalancing cannot always be used. Given that the number of all possible combinations is a permutation of the number of conditions, or k!, the number of possible sequences gets larger as the number of conditions increases. With 7 conditions, 7! or 5,040 different orders should be used to obtain a complete counterbalanced design. Therefore, the use of complete counterbalancing is generally recommended only for k ≤ 4. When k > 4, the researcher will usually settle for incomplete counterbalancing. Two incomplete designs are considered: a balanced square and a Latin Square.

When a researcher expects sequence effects, he or she may employ a balanced square design, which requires that (a) each condition occurs an equal number of times in each ordinal position, and (b) each condition is preceded by each of the other conditions an equal number of times. When the number of conditions is even, only a set of k sequences are needed; if the number of conditions is odd, two sets of k sequences are needed, with the second mirroring exactly the first. If first-order carry-over effects (from one condition to the one immediately following it) are present, this design affords the same control as complete counterbalancing.

When the researcher's focus is on controlling order effects (e.g., fatigue), a Latin square can be used. Table 2 shows a possible Latin squares for the computer-mediated communication study of the previous example, now comprised of four conditions (video, chat, text-message, and face-to-face).

Table 1 Example of Complete Counterbalancing for a CMC Experiment With Three Conditions

Table 1 Example of Complete Counterbalancing for a CMC Experiment With Three Conditions

	Time/Position 1	*Time/Position 2*	*Time/Position 3*
Subject/Group 1	Self-Efficacy A	Self-Efficacy B	Self-Efficacy C
Subject/Group 2	Self-Efficacy B	Self-Efficacy C	Self-Efficacy A
Subject/Group 3	Self-Efficacy C	Self-Efficacy A	Self-Efficacy B
Subject/Group 4	Self-Efficacy A	Self-Efficacy C	Self-Efficacy B
Subject/Group 5	Self-Efficacy B	Self-Efficacy A	Self-Efficacy C
Subject/Group 6	Self-Efficacy C	Self-Efficacy B	Self-Efficacy A

Independent variable: type of CMC (*A* = video, *B* = chat, *C* = text message)
Dependent variable: self-efficacy

Table 2 Example of a Latin Square Design for a Computer-Mediated Communication Experiment With Four Conditions

Table 2 Example of a Latin Square Design for a Computer-Mediated Communication Experiment With Four Conditions

	Time/Position 1	*Time/Position 2*	*Time/Position 3*	*Time/Position 4*
Subject/Group 1	Self-Efficacy A	Self-Efficacy C	Self-Efficacy B	Self-Efficacy D
Subject/Group 2	Self-Efficacy B	Self-Efficacy A	Self-Efficacy D	Self-Efficacy C
Subject/Group 3	Self-Efficacy C	Self-Efficacy D	Self-Efficacy A	Self-Efficacy B
Subject/Group 4	Self-Efficacy D	Self-Efficacy B	Self-Efficacy C	Self-Efficacy A

Independent variable: type of CMC (A = video, B = chat, C = text message, D = face to face)
Dependent variable: self-efficacy

Three requirements characterize a Latin square design: (1) Each condition occurs equally often in each ordinal position (1st, 2nd, 3rd); (2) each condition occurs only once in each row and in each column; and (3) the number of rows (participants) equals the number of columns (conditions), such that there must be at least as many subjects as conditions, and the number of participants must be a multiple of k. Thus, in a five-condition experiment there must thus be at least 5 participants (1 per group) or a multiple of 5 (10, 20, 25, and so on).

While a Latin square affords control over order effects, it must be noted that what appears like a main effect of any of the three components (row, column, or condition) can always be due to a more complex interaction between treatment and order or sequence. When specifying a counterbalanced design, the researcher should thus consider whether an interaction is more or less likely based on theoretical and empirical considerations. If it is unlikely, the researcher can proceed and subsequently confirm the absence of interaction effects through analysis of variance (ANOVA) by testing for a sequence main effect, or a treatment by ordinal position interaction. However, if the presence of a significant interaction could reasonably be expected that would substantially undermine the interpretation of the experiment, the researcher should consider redesigning the study.

When even incomplete counterbalancing is not feasible, randomized counterbalancing can be applied by exposing each participant to a randomly determined ordering of conditions. Although this method ensures that order effects and sequence effects between adjacent trials will be balanced across the different randomizations, sequence effects within any given participant will be most likely unbalanced. Successful counterbalancing through randomization can be assessed post-hoc; however, randomized counterbalancing does not allow the researcher to test for the presence of interactions between the procedural or blocking variables and the experimental variables.

Application of the Latin Square to Counterbalancing

When creating a Latin square counterbalanced design, two options are available: (1) The same Latin Square can be used multiple times, or (2) different Latin squares can be used. In both cases, the Latin squares can be constructed or selected randomly from the population of all possible squares available for a given number of conditions (for k = 4,576 different squares are possible).

Imagine a five-condition experiment. When the same Latin square is used, Subjects 1–5 would be assigned respectively to the 1st, 2nd, 3rd, 4th and 5th row; Participant 6 would be then assigned to the same row as Participant 1, Participant 7 to the same row as Participant 2, and so. However, due to chance, it is possible that order or sequence interacts with conditions in the particular Latin square chosen by the researcher, thus making the main effects of condition uninterpretable. Carry-over effects may be especially likely, given that it is not a Latin square requirement that each condition is preceded by every other one equally often. However, the presence of such interactions can be assessed by performing a test of square uniqueness. If significant, the test indicates that incomplete counterbalancing was not sufficient.

To decrease the chances that confounding interactions within a single Latin square inflate the error estimates, several different Latin squares, selected randomly, can be adopted in the same design and applied to successive groups of participants. This choice allows greater control over carry-over effects, especially if the design is expanded to a balanced Latin square, which is the optimal solution if enough participants are available. However, no significance test is available to assess whether partial counterbalancing was sufficient when multiple Latin squares or balanced Latin squares are used.

Elena F. Corriero

See also Blocking Variable; Internal Validity; Latin Square Design; Repeated Measures; Within-Subjects Design

Further Readings

Campbell, D. T., & Stanley, J. C. (1967). Experimental and quasi-experimental design for research. Boston, MA: Houghton Mifflin. doi:10.1037/022808

Gaito, J. (1961). Repeated measurements designs and counterbalancing. Psychological Bulletin, 58(1), 46–54. doi:10.1037/h0046169

Gould, J. E. (2001). Concise handbook of experimental methods for the behavioral and biological sciences. Boca Raton, FL: CRC Press.

Kirk, R. E. (1995). Experimental design: Procedures for behavioral sciences (3rd ed.). Pacific Grove, CA: Brooks/Cole.

Elena F. Corriero

http://dx.doi.org/10.4135/9781483381411.n103

10.4135/9781483381411.n103