



## A Revision of the Campbellian Validity System

Thorleif Lund

To cite this article: Thorleif Lund (2020): A Revision of the Campbellian Validity System, Scandinavian Journal of Educational Research, DOI: [10.1080/00313831.2020.1739126](https://doi.org/10.1080/00313831.2020.1739126)

To link to this article: <https://doi.org/10.1080/00313831.2020.1739126>



© 2020 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 19 Mar 2020.



Submit your article to this journal [↗](#)



Article views: 95



View related articles [↗](#)



View Crossmark data [↗](#)

# A Revision of the Campbellian Validity System

Thorleif Lund

Department of Special Needs Education, University of Oslo, Norway

## ABSTRACT

The purpose of this paper is to propose a revision of the well-known Campbellian system for causal research. The revised system, termed the COPS model, applies to both applied and basic research. Five validities are included, where two validities are adopted from the Campbellian system, and the validities are partly hierarchically ordered. Important similarities and differences between the Campbellian system and the COPS model are discussed.

## ARTICLE HISTORY

Received 7 March 2019  
Accepted 16 January 2020

## KEYWORDS

Campbellian validity system;  
quality criteria; causal  
research; COPS model

Knowledge construction requires a set of criteria for evaluating the quality of research-based knowledge. Such a set of quality criteria – often termed a validity system – should be well-defined and consistent, pragmatic and relevant to various kinds of basic and applied research, and justified philosophically.

Campbell and co-workers (Campbell & Stanley, 1963; Cook & Campbell, 1979; Shadish et al., 2002) have proposed a validity system for causal research, and this system has been used extensively in quantitative social science (Shadish, 2010). The system has been discussed by several methodologists (e.g., Cronbach, 1982; Guba & Lincoln, 1982; Judd & Kenny, 1981; Kruglanski & Kroy, 1976; Lund, 2010; Mark, 1986; Mook, 1983; Reichardt, 2002, 2006, 2008).

The purpose of the present paper is to propose a revision of the Campbellian system, termed the COPS model. The account below is divided into three main sections. A short description of the Campbellian system is given in the first section, the COPS model is presented in the second section, whereas the system and the model are compared with respect to important themes in the final section. The focus is on conceptual matters, omitting statistical and technical details.

Several themes of relevance to the Campbellian system are not discussed in the present paper, e.g., alternative concepts of construct validity, alternative definitions of critical realism, alternative methodological/statistical effect definitions, and alternative validity systems (except for a brief account of Cronbach's system in the final section). Even though such themes are important ones, a satisfactory treatment of them would require considerably more space than available here.

## The Campbellian Validity System

The validity system was first presented in Campbell and Stanley (1963), and later revised by Cook and Campbell (1979) and by Shadish et al. (2002). The original system contains two kinds of validity, whereas four validities are defined in the two revisions. Since the latter revision is presumably intended by its authors to replace its predecessors, the present account is related to the system in

**CONTACT** Thorleif Lund  thorllun@broadpark.no  Korsveien 2, 1455 Nordre Frogn, Norway

© 2020 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group  
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

Shadish et al. (2002), which – for simplicity – is abbreviated as SCC. Below, central aspects of this system are briefly described, and these aspects will be elaborated in later sections.

The 2002 system consists of four inferences, their respective validities, and a set of possible threats against each validity. The four inferences are *statistical*, *causal*, and *construct* inferences – i.e., inferences from observations to constructs – as well as *generalizations* of causal relationships. The validities of these inferences are *statistical conclusion*, *internal*, *construct*, and *external* validity, respectively. The term *validity* refers to the approximate truth of inferences, and validity is therefore a property of inferences, not of methods, data, or results. Validity may vary in degree since the validity in a given case depends on the extent relevant evidence supports the inference in question as being true. The system is based on the philosophical assumption of critical realism, and on a fallibilist version of falsification of how to obtain validity of inferences (Cook & Campbell, 1979, chapter 1). The system is in line with an experimentalist view on knowledge construction, and is therefore especially relevant for experimental and quasi-experimental research, but it is meant to apply also to non-experimental studies.

The four validities are defined in SCC as follows: (a) Statistical conclusion validity concerns the truth of “two related statistical inferences that affect the covariation component of causal inferences: (1) whether the presumed cause and effect covary and (2) how strongly they covary” (p. 42). Thus, two statistical criteria are involved with this validity, one relating to statistical tests and confidence intervals of the covariation, and the other to effect size. (b) Internal validity is the truth of “inferences about whether observed covariation between A and B reflects a causal relationship from A to B in the form in which the variables were manipulated or measured” (p. 53). (c) Construct validity is defined as “the degree to which inferences are warranted from the observed persons, settings, and cause and effect operations included in a study to the constructs that these instances might represent” (p. 38). (d) External validity is defined as “the validity of inferences about whether the causal relationship holds over variation in persons, settings, treatment variables, and measurement variables” (p. 38).

Hence, the validity system comprises four kinds of inferences, and the four validities stand for the (approximate) truth of these inferences. Internal validity is the central validity in the system. Threats against validity are reasons why inferences can be incorrect, and the validity is satisfactory to the degree that such threats can be considered falsified by design control, statistical control, or by rational/theoretical arguments. The strategy for estimating the degree of validity of an inference is therefore based on a falsificationistic perspective.

The description above shows that statistical conclusion and internal validity are concerned with study operations, and with the relationship between treatment and outcome. Statistical conclusion validity depends on falsification of errors in assessing statistical covariation, while internal validity depends on falsification of causal-reasoning errors.

To elaborate the concept of internal validity, Campbell (1986) has relabelled this validity as *local molar causal validity*, and the label is explained by SCC (p. 54) as:

The word *causal* in *local molar causal validity* emphasizes that internal validity is about causal inferences, not about other types of inferences that social scientists make. The word *local* emphasizes that causal conclusions are limited to the context of the particular treatments, outcomes, times, settings, and persons studied. The word *molar* recognizes that experiments test treatments that are a complex package consisting of many components, all of which are tested as a whole within the treatment condition. ... Of course, experiments can and should break down such molar packages into molecular parts that can be tested individually or against each other. But even those molecular parts are packages consisting of many components.

Hence, internal validity is defined at the level of sampling particulars, and is therefore independent of constructs and construct validity, and it concerns the impact of complex packages of components. It should be added that the internal-validity concept is in line with causal description (SCC, p. 98).

Construct and external validity concern generalization, the former generalizations from sampling particulars to constructs, and the latter generalizations of internal-validity causal relationships across variations in persons, settings, treatments, and outcomes. A problem of construct validity is a representation problem, since it deals with which constructs are represented by the sampling particulars. A construct is a named, specified class of elements, i.e., a construct encompasses a construct name,

construct criteria, and a class of elements as specified by the construct criteria. Construct validity is therefore adequate to the extent the criteria for selecting sampling particulars match the construct criteria, such that the particulars can be interpreted as a representative subset of the chosen construct class. Construct validity is independent of causal relationships. External validity, on the other hand, is related to the involved causal relationships, and this validity is independent of constructs and construct validity (SCC, pp. 93–95). It should further be noted that external-validity generalizations are meant to have broad application. For example, they concern variations in persons, settings, treatments, and outcomes that were in the study as well as variations in persons, settings, treatments, and outcomes that were not in the study, and several types of generalizations are included: from narrow to broad, from broad to narrow, at a similar level, to a similar or different kind, and from a random sample to its population (SCC, pp. 83–84).

Time is not included in the definitions of construct and external validity, but these validities are relevant for time, as well. According to SCC (p. 20): “We occasionally refer to time as a separate feature of experiments, following Campbell (1957) and Cook and Campbell (1979) because time can cut across the other factors independently.”

Even though the Campbellian system has a dominant position in causal research, parts of the system have been criticized. The criticism concerns mainly internal validity, e.g., that this validity is independent of constructs. Several writers (e.g., Cronbach, 1982; Kruglanski & Kroy, 1976; Lund, 2010; Reichardt, 2008) have argued that the defined independence of internal and construct validity is not meaningful, and that a causal inference should not be limited to sampling particulars, but should include constructs, as well. Another important criticism is Mook’s (1983) argument that the Campbellian external validity is not obligatory for basic research, because the purpose of such research is not to generalize across persons, settings, treatments, or outcomes, but to generate theoretical interpretations.

The revision of the Campbellian system presented below avoids the mentioned criticisms, and will improve the system also in other respects.

## An Alternative to the Campbellian Validity System

The revised system is a model for inferences and their validities in applied and basic quantitative causal research, and is termed the *COPS model* (an abbreviation of the cause-, outcome-, persons-, setting-model). The model is meant to be relevant for experimental, quasi-experimental, and non-experimental studies. The term “causal relationship” is used synonymously with “causal effect” or simply “effect”, and the term “sampling particular” is often replaced by the shorter term “indicator”.

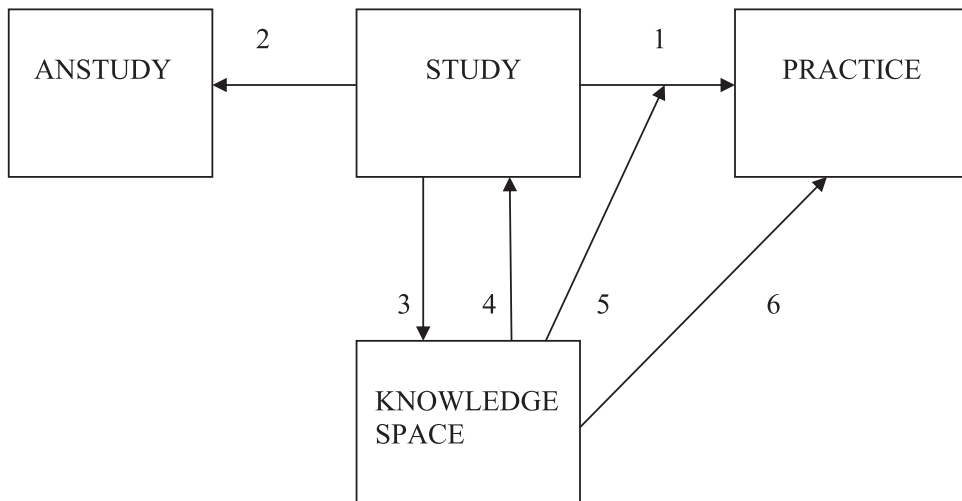
In short, the COPS model can be characterized as follows: The inferences and their validities involve constructs and are conceptualized within a broad context. In this context, generalization is defined widely as the transfer of “some information” from “some place” to “another place”. The model contains five validities, their respective inferences, and a set of threats against each validity, where the validities and their inferences are partly hierarchically ordered. As in the system, validity is the approximate truth of an inference. The model is based on the same philosophical assumption of critical realism and on the same falsificationistic perspective as the system. The four features of treatment, outcome, persons, and setting in the Campbellian system are retained in the model, apart from that the term “treatment” is replaced by the term “cause” to indicate that a causal factor in the model refers to a more comprehensive class of causes than treatments. A cause in the model may be a treatment, a treatment combination, a treatment component, a factor in a natural context, or an organismic property. The concept of cause adopted here is probabilistic and broad so as to encompass intention and purpose (Cook & Campbell, 1979, chapter 1), and the concept stresses the productive (dynamic, active) character of causation (Kenny, 1979, p. 4). The model is also relevant for time, even though time is not defined as a feature together with the four other ones, which is in line with the account of SCC (p. 20).

The five inferences in the COPS model are *statistical, construct, causal, generalization, and theoretical* inference, and their respective validities are *statistical conclusion, construct, causal, generalization, and theoretical* validity. The two former inferences and their validities are adopted unchanged from the Campbellian system, whereas a causal inference and its causal validity in the model deviate from a causal inference and its internal validity in the system, a generalization inference and its generalization validity in the model deviate from a generalization and its external validity in the system, and a theoretical inference and its theoretical validity are categories which cannot be compared to inferences/validities in the system. As for the former four validities, statistical conclusion and construct validity do not depend on each other or on the two other ones, but causal validity depends on statistical conclusion and construct validity, and generalization validity depends on causal validity. The same pattern holds for the four respective inferences. Hence, the former four validities and their inferences are partly hierarchically ordered, with statistical conclusion and construct validity/inference at the bottom, with generalization validity/inference at the top, and with causal validity/inference in between. How theoretical validity and inference are defined and related to the other validities/inferences are accounted for later.

As for validity threats, the threats against the Campbellian validities are relevant with the former four COPS validities, as well, but the threats are combined in line with the partly hierarchical order. The threats against statistical conclusion and construct validity are the same as before. However, the threats against causal validity are the combination of the threats against the Campbellian statistical conclusion, construct, and internal validity, while the threats against generalization validity are the combination of the threats against all four Campbellian validities. In line with the falsificationistic perspective, each validity is satisfactory to the extent its respective threats are falsified.

The artificial therapy study sketched below and [Figure 1](#) are used for definition and illustration of the validities/inferences in the COPS model, as well as for the comparison of the COPS model and the Campbellian system.

Suppose a new variant of individual cognitive therapy has been proposed as a supplement to an ordinary program for the treatment of juvenile delinquents. The therapy focuses on such adolescents' conflict level, i.e., on reducing frequency and intensity of their conflicts with others. Before the therapy is implemented in professional work, its effect has to be assessed. The research problem in a therapy study is whether the therapy has a considerable impact on the conflict level for such adolescents in a clinical setting. Let the methods and results in the study be as follows: A large,



**Figure 1.** Illustration related to the COPS model (see text).

representative sample of male and female juvenile delinquents is randomly distributed to three groups, i.e., a therapy group (T-group), an attention-control group (C<sub>1</sub>-group), and a control group with the ordinary program only (C<sub>2</sub>-group). The therapy in the T-group comprises daily sessions during eight weeks, and is given by therapists in a clinical setting, whereas the C<sub>1</sub>-group is exposed to a treatment assumed to correspond to general attention in the T-group. Except for the omitted specific therapy component, the C<sub>1</sub>-condition contains the same components as the T-condition. The T- and the C<sub>1</sub>-group receive in addition the ordinary program. Immediately before and after the treatment interval, each adolescent is given a score on a rating scale meant to represent conflict level, i.e., from 0 (“no conflicts”) to 100 (“extremely high conflict level”), and the conflict scores are based on individual interviews with the adolescents and their parents as well as on other data (e.g., school and police records). Hence, a randomized pretest–posttest design with two control groups is used. The mean change (posttest score – pretest score) is –35 for the T-group, –15 for the C<sub>1</sub>-group, and –5 for the C<sub>2</sub>-group. Consequently, the T/C<sub>1</sub>-difference (–35 – (–15)) is –20, the T/C<sub>2</sub>-difference (–35 – (–5)) is –30, and the C<sub>1</sub>/C<sub>2</sub>-difference (–15 – (–5)) is –10. The three differences are regarded as of considerable size and are statistically significant.

Two causal effects are relevant for solving the stated research problem in the therapy study above: one related to the impact of the total treatment package in the T-condition, and another related to the specific therapy component. The two quantitative effect estimates are –30 and –20, respectively. In addition, the effect of general attention can be measured, and the quantitative estimate is here –10. Note that statistical conclusion validity is satisfactory in all three cases. Construct validity is satisfied to the extent the indicators in the study are representative for their respective construct classes. Causal and generalization validity in connection with the study are accounted for later. The therapy example can be considered typical applied research, i.e., a kind of evaluation research (Shadish et al., 1991).

Figure 1 encompasses a research study (here termed study), another research study (termed anstudy), a situation which is not a research study (termed practice), a set of knowledge and quality criteria (termed knowledge space), and six arrows (numbered 1–6) representing generalizations.

The elements in Figure 1 can be briefly specified as follows: By study is meant an individual causal research study, as illustrated by our therapy study. A similar interpretation is given to anstudy which has been accomplished at the same time as study or earlier, and anstudy is included in the diagram only for illustrating the comparison of study and another study with respect to measured causal relationships. Furthermore, practice is a situation without measured causal relationships, and to which causal study results are generalized, e.g., some professional situation related to treatment of juvenile delinquents. It is assumed that study, anstudy, and practice differ with respect to at least one of the cause (C)-, outcome (O)-, person (P)-, and setting (S)-construct classes. For simplicity, such classes are omitted in the diagram. In line with Lund (2005, 2013), knowledge space is defined as the set of substantive and methodological knowledge within a research domain provided by earlier research as well as of research standards, e.g., validity systems and ethical criteria. The substantive and methodological knowledge may be in terms of theories, and substantive knowledge can be basic or applied knowledge. In connection with our therapy example, knowledge generated by earlier research related to treatment of juvenile delinquents constitutes essential parts of knowledge space. Philosophy-of-science assumptions are not included in knowledge space, but can be considered justifications of some parts of knowledge space, for example of validity systems.

The arrows 1–6 stand for six main kinds of generalizations in line with the wide definition of generalization given above. Arrows 1 and 2 stand for transfer of study results to practice and anstudy, respectively, whereas arrow 3 represents transfer of study findings and inferences/validities to knowledge space. The study findings involved in these transfers may be causal or non-causal ones, but the focus here is on causal results. By arrow 4 is meant transfer to study of knowledge and standards in knowledge space, e.g., for the purpose of assisting measurements in study. Arrow 5 stands for generalization of knowledge in knowledge space in order to support or moderate 1-generalizations. By arrow 6 is meant estimation of a causal relationship in practice by transferring information from

knowledge space. 6-generalizations are useful only before a study is undertaken, not afterwards because then the direct path from study to practice – i.e., the 1-generalizations – will yield the same practice results as the indirect path from study to practice via knowledge space – i.e., the combination of 3- and 6-generalizations. The 6-generalizations are therefore not required in the figure since study is included, but they are nevertheless presented in order to illustrate the dependence of professional work on generalizations from earlier research.

Even though [Figure 1](#) contains only one practice, several such targets may be included. Moreover, study may enter a larger research program as part of a set of similar individual studies, as in a meta-analysis of individual study results (Hunter & Schmidt, 1990; Smith & Glass, 1977). For simplicity, such additional targets and individual studies are omitted in the diagram, except for an study.

The COPS model applies to both applied and basic research. All 1- to 5-generalizations in [Figure 1](#) are involved in applied research, but often only 3- and 4-generalizations in basic research. Below, inferences and validities related to applied research are accounted for first, followed by a short description of how the COPS model applies to basic research. The section of applied research includes definitions of causal validity/inference and generalization validity/inference, whereas theoretical validity/inference are defined in connection with basic research. The account is limited to quantitative causal research.

## **The COPS Model and Applied Research**

### **Effect Measurement**

Effect measurement in the COPS model requires that causal effect, causal inference, and causal validity are defined at the construct level, not at the indicator level. Causal validity in study is defined as the truth of the causal inference about the causal relationship between a representative cause in a causal construct class (C) and a representative outcome in an outcome construct class (O) for a representative person in a person construct class (P) and for a representative setting in a setting construct class (S). Or, put otherwise and somewhat imprecisely: ... truth of the inference that a causal construct class has produced a change in an outcome class for a person construct class and for a setting construct class ... . Causal validity depends on statistical conclusion and construct validity, and on the design in study, and is also influenced by knowledge-space information in terms of 4-generalizations. Causal validity is satisfactory to the extent the actual statistical conclusion and construct validities are high, and requires also that the three ordinary criteria of causality are satisfied, i.e., that the presumed cause precedes the presumed effect, that they covary, and that irrelevant factors are controlled for. If alternative causes are controlled for statistically, e.g., in quasi- and non-experimental research, these causes presuppose constructs.

Hence, a COPS effect is a change in outcome produced by a cause, where the cause and the outcome are defined at the construct level. The COPS effect can be a short-term or long-term effect. Each person in the person construct class has such a produced change, and the COPS effect for the person construct class is the mean of such individual changes. Note, moreover, that although the effect depends on which person and setting constructs are involved, these constructs are not considered causes since they lack the producing quality.

If causal inferences have satisfactory validity, they will be published, i.e., transferred to knowledge space in terms of 3-generalizations. It follows that only 3- and 4-generalizations in [Figure 1](#) are involved with effect measurement in the COPS model.

Our therapy study can be used for illustration. –30 is an estimate of the COPS effect of the therapy-package construct class on the conflict construct class for the juvenile delinquents construct class and for the clinical setting construct class. This COPS effect is, therefore, a change in conflict at the construct level during the pretest–posttest interval as produced by the therapy-package at the construct level. The causal inference is the statement about this measurement, and causal validity is the truth of this causal inference. Causal validity is satisfactory to the extent that statistical conclusion validity is satisfactory, that the involved constructs are correctly specified, and that actual

threats related to the Campbellian internal validity are falsified. Thus, causal validity depends on statistical conclusion and construct validity, and the threats against causal validity are a combination of the threats against the Campbellian statistical conclusion, construct, and internal validity.

Knowledge space will influence the therapy study in terms of 4-generalizations as follows: The general aim and research problem in the therapy study are chosen in line with the needed information as suggested by the scrutiny of knowledge space. Moreover, 4-generalizations will affect statistical conclusion, construct, and causal validity. Relevant statistical-theory information in knowledge space is transferred to study in terms of 4-generalizations, e.g., standards concerning significance level, statistical power, and effect size. Construct validity may be influenced by knowledge-space information, as well, e.g., in connection with the conflict construct. Suppose the conflict measure has been used in earlier research. If so, the actual construct validity in the therapy study may be transferred, totally or partly, from knowledge space. Since causal validity depends on statistical conclusion and construct validity, also causal validity will be influenced by such 4-generalizations. Furthermore, according to design theory in knowledge space, the randomized assignment of the adolescents to the three conditions will strengthen causal validity. Suppose that a simple pretest–posttest design had been used instead, that the therapy effect had been estimated by the pretest–posttest change, and that earlier research suggests that such adolescents without therapy will show no change in conflict. This information can be transferred by a 4-generalization and then increase the causal validity in our therapy study. If, on the other hand, the knowledge-space information suggests a conflict reduction, this reduction can be subtracted from the pretest–posttest change, thus modifying the effect estimate in study.

The preceding account illustrates that effect measurement – and the related causal inference and validity – will depend not only on information within study, but also on generalization from knowledge space in terms of 4-generalizations. Hence, in this sense, effect measurement is a function of generalization, as well.

The COPS model encompasses effect measurement also for subclasses or components within C-, O-, P-, and S-construct classes. For example, whereas –30 is the estimated effect of the total therapy package in our study, –20 is the effect estimate of the therapy component within this treatment package class. Analogously, we may be interested in an effect estimate for an outcome subclass (e.g., intensity of conflict), for a person subclass (e.g., male juvenile delinquents), or for a setting subclass (e.g., a particular component in the clinical setting). Such effect measurement for subclasses is accounted for as above, apart from that causal effect, causal inference, and causal validity will now be defined for the actual construct subclasses.

Above, effect measurement is elaborated for study in [Figure 1](#), but a similar account can be given for anstudy.

### ***Effect Generalization***

Applied-research studies include not only causal inferences/validities, but also generalization inferences/validities. Generalization validity is defined as the truth of the generalization inference that a COPS effect holds over studied or unstudied variation in C-, O-, P-, or S-construct classes or subclasses. By studied variation is meant that the involved COPS effects are measured, whereas only the study effect is measured in an unstudied variation. The two kinds of variation are illustrated in [Figure 1](#) by arrows 2 and 1, respectively. While the COPS effects in study and anstudy are measured, the unmeasured COPS effect in practice is estimated by generalization from study and knowledge space as illustrated by arrows 1 and 5. A generalization inference and its validity require, of course, that the involved effects in the variation of interest are defined. The COPS effect in study and anstudy is defined as above, whereas a practice effect will be defined in an analogous way on the basis of a scrutiny of practice. If a generalization inference has satisfactory generalization validity, it will be published, i.e., transferred to knowledge space in terms of a 3-generalization. Since generalization validity depends on causal validity, such a generalization inference requires that also the involved causal inferences are sufficiently valid.



Studied variation can be divided into variation between and within studies, and in both cases the involved measured COPS effects are compared. The between-variation case is illustrated by arrow 2 in Figure 1, and generalization validity with this 2-generalization depends on the similarity between study and anstudy effects, and on causal validity in connection with these two measured COPS effects. Suppose that study in the diagram represents our therapy study using cognitive therapy, that anstudy stands for a therapy study using psychodynamic therapy and is similar to our therapy study with respect to outcome, persons, and setting, and that the two therapies produce the same therapy-package effects. The COPS effect can then be generalized across the two therapy construct classes. A similar reasoning holds for a within-study variation (not illustrated in Figure 1), e.g., high similarity between the package effects for boys and girls in our therapy example implies that the same measured COPS effect can be generalized across the two gender subclasses.

As for unstudied variation, generalization validity depends on causal validity in connection with the study effect, on the general similarity between study and practice, and on knowledge-space information. The knowledge-space information may support or reduce generalization validity. If an earlier research study is somewhat similar to practice and has given a similar causal result as study, this information in terms of a 5-generalization may strengthen generalization validity. On the other hand, if this earlier study has produced a quite different causal effect, the generalization validity may be reduced.

Generalization inferences from study to practice can be divided into the case where the causal factor in the inference exists in practice and the case where practice lacks this factor. Hence, the former inference is relevant in situations where we estimate the effect of an established causal factor in practice from a research study, while the latter inference takes place when a causal factor in study is supposed to be added to a practical setting without this factor. Both kinds of inferences are relevant in educational and psychological research. The latter inference is illustrated by our running example if the estimate effect of therapy is transferred to a professional situation without this therapy. Generalization validity will be estimated by the same strategy for the two kinds of generalization inferences, i.e., by comparing study and practice features in combination with possible knowledge-space information in terms of 5-generalizations.

Generalization validity encompasses also component effects, not only package effects. For instance, the therapy-component effect of  $-20$  in our therapy example may be transferred to a professional situation in terms of a 1-generalization. Component-effect generalizations within treatment packages may also be relevant with studied variation.

It should be stressed that generalization validity is defined at the construct level, since this validity concerns constructs related to measured or unmeasured COPS effects in the involved studied or unstudied variations. For example, a therapy effect of  $-30$  in study in the 1-generalization is given for specified constructs of therapy, conflict, adolescents, and clinical setting, and analogue specifications are required for the practice effect. A similar account holds for studied variation as illustrated by the 2-generalization in Figure 1. In addition, if unwanted variation is controlled for, also this variation is described in terms of constructs.

Generalization validity presupposes that study, anstudy, and practice are different with respect to at least one of the C-, O-, P-, and S-construct classes. If not, we have cases of effect measurement instead of effect generalization. That is, if study and anstudy include the same constructs, the study and anstudy effects at the construct level are identical, and the two measured effects will therefore estimate the same defined effect. Similarly, identical constructs in study and practice will imply the same effect at the construct level in study and practice. It should further be noticed that since generalization validity depends on causal validity, and since no validity is perfect, generalization validity will be lower than the involved causal validity (or validities) in an empirical case.

In line with SCC (p. 20), even though effect generalization in the COPS model is relevant also across time, time is not included explicitly in the model together with cause, outcome, persons, and setting. The reasons for this omission are that time is an ambiguous concept in that it can refer either to historical occasions or to a feature in an empirical investigation, and that time runs

across the other four features. Thus, a COPS effect – a short-term or long-term effect – is defined as a change in outcome in a time interval, a COPS cause operates in a time interval, and persons and setting are located in intervals. Time is therefore contained implicitly in the model. The reasons above do not mean that time is less important than the other four features. For example, in our therapy study, if the package effect of  $-30$  is interpreted as a short-term effect in the pretest–posttest interval, the critical question is whether this effect holds over time. Lack of follow-up measurements should therefore be considered a serious weakness of the design used, and estimation of a long-term effect in this case depends on knowledge-space information in terms of a 5-generalization in [Figure 1](#). If such information is lacking, estimation will be pure guesswork.

By study in [Figure 1](#) is meant an individual investigation, but the study could alternatively be a meta-analytic investigation (Hunter & Schmidt, 1990; Smith & Glass, 1977), e.g., a synthesis of a set of individual therapy studies. Such a synthesis can be accounted for in analogy with the reasoning above.

### **The COPS Model and Basic Research**

The primary purpose of basic research is to develop theory, not to generalize findings to or across “real-life” situations as in applied research. Both theory-testing and exploratory approaches can be used for this purpose. In the former case, a prediction/hypothesis is derived from a theory and tested in an empirical study, and the theory is supported, unsupported, or modified on the basis of the findings. In the exploratory case, theoretical statements are generated in the study without starting with a theory. An exploratory case will ordinarily be followed by theory-testing studies.

Mook (1983) has pointed out that the Campbellian external-validity concept is not obligatory for basic research, since such research can result in valuable theoretical conclusions even though the conclusions are externally invalid. However, as also mentioned by Mook, basic research will sometimes provide findings which can be generalized in line with Campbellian external validity.

Below, it is briefly shown how the COPS model gives an adequate representation also of basic research, not only of applied research. The account assumes causal research, is limited to the theory-testing case, and [Figure 1](#) is used for illustration.

In short, the theory-testing case is accounted for by the COPS model as follows: (a) A theory located in knowledge space is transferred to study by a 4-generalization, (b) in study, a construct-involved causal hypothesis is derived from the theory and tested on the basis of the study findings, which results in a causal inference, (c) the theory is supported, unsupported, or modified, depending on whether the hypothesis is verified or unsupported, and (d) the resulting theoretical conclusion is transferred back to knowledge space by a 3-generalization. This construct-included conclusion is here termed a *theoretical inference*, and *theoretical validity* is the (approximate) truth of this inference. Note that 1-, 2-, and 5-generalizations, which concern generalization inference/validity, are not involved in this theory-testing strategy, only 3- and 4-generalizations.

Hence, according to the COPS model, test of the theory depends on the test of the derived hypothesis, and theoretical inference/validity will depend on causal inference/validity as defined earlier. If the hypothesis is supported, also the theory is supported, and vice versa. In this theory-testing approach, theoretical inference/validity will be of primary interest, whereas causal inference/validity should be considered necessary but subordinate means for generating theoretical inference/validity.

The preceding account can be illustrated by Harlow’s (1958) famous experiment of “mother love” in young rhesus monkeys, which is used by Mook (1983) in his critical article. The purpose of the study was to test the drive-reduction theory. Harlow’s independent variable consisted of two conditions: wire “mothers” and terry-cloth “mothers”, where the wire “mothers” were supplied with food. The dependent variable was the amount of attachment the monkeys showed in the two conditions. In line with the drive-reduction theory it was predicted that the monkeys should prefer the wire mothers, but the monkeys chose mainly the terry-cloth mothers. A reasonable interpretation was that contact comfort was a powerful determinant of attachment, whereas nutrition played a

minor role. The conclusion was therefore that the theory should be considered unsupported or at least questionable and in need of modification.

Mook points out that the Campbellian external-validity criterion is clearly violated here, because the two experimental conditions, the sample of monkeys, and the experimental setting are artificial, i.e., extremely unrepresentative of any natural “real-life” contexts with young monkeys. The causal findings can therefore not be generalized validly to such contexts. However, as stressed by Mook, the purpose of Harlow was not to make such generalizations, but to test the drive-reduction (or hunger-reduction) theory, and the external-validity concept is consequently irrelevant in this study. Mook adds that Harlow’s theoretical conclusion can be considered a kind of “understanding” of the underlying process, and that if anything can be generalized externally it is this “understanding”. Note, however, that such an “understanding” generalization is quite different from the Campbellian external-validity effect generalizations.

Harlow’s study can be accounted for by the COPS model in line with the four points above: (a) The drive-reduction theory in knowledge space is transferred to study by a 4-generalization, (b) a causal hypothesis is derived from the theory and tested in study, which (c) results in that the theory is unsupported, and (d) the negative conclusion about the theory is transferred back to knowledge space by a 3-generalization. This conclusion and its (approximate) truth are termed theoretical inference and theoretical validity, respectively.

From a broad theory such as the drive-reduction theory several kinds of causal hypotheses can be derived, which require many studies. In each study, the sampling particulars should be chosen on the basis of the C-, O-, P-, and S-constructs in the actual hypothesis, so that the sampling particulars can be considered representative of their respective construct classes. For example, in Harlow’s study, also the constructs in the relevant hypothesis will be artificial, not only the sampling particulars.

As for theoretical validity in a given study, this validity will depend on causal validity, but also on how relevant the derived hypothesis is in relation to the theory, on the *a priori* validity of the theory, and on how broad the theory is. For example, the broader a theory, the smaller the impact of an isolated study on theory tends to be. Consequently, satisfactory testing and developing a broad theory require ordinarily a large set of individual studies. Finally, it follows that the threats against theoretical validity consist of causal-validity threats as well as logical/rational threats related to the derivation of the hypothesis.

The brief account above illustrates how the COPS model applies to the theory-testing approach of basic research, but the model is also relevant for an exploratory case of such research. In short, according to the model’s conceptualization of the latter case, a theory is generated in study on the basis of study findings and knowledge-space information, and the theory is generalized to knowledge space. As before, this generalization is a theoretical inference, and its truth is theoretical validity.

## Final Discussion

The preceding account shows that the COPS model deviates from the Campbellian system in the following important ways: (1) Statistical conclusion, construct, causal, and generalization validity in the model are defined in a partly hierarchical order, whereas the four Campbellian validities are independently defined. Statistical conclusion and construct validity in the system are considered useful criteria, and are therefore included in the model, but the Campbellian internal validity and the external validity are substituted by causal validity and generalization validity, respectively, in the model. Moreover, theoretical validity depends on causal validity. (2) Causal validity, generalization validity, and theoretical validity in the model are defined at the construct level, whereas internal validity and external validity in the system are defined at the indicator level. Hence, these definitions in the model are in line with the arguments by the critics of the system that causal inferences and validities should include constructs (e.g., Cronbach, 1982; Kruglanski & Kroy, 1976; Lund, 2010; Reichardt, 2008). (3) Since causal validity depends on statistical conclusion and construct validity, and since no validity will be perfect, causal validity is lower than internal validity in an empirical case. A similar reasoning

holds for generalization validity compared to external validity, i.e., the former will be lower than the latter. (4) Generalizations in the model are defined within a broad context, where the concept of knowledge space plays a central role, while such a concept is lacking in the system.

Knowledge space is a useful concept in accounting for effect measurement and generalization in applied research. Another important advantage is that the concept makes possible that the COPS model takes care of Mook's (1983) argument that the Campbellian external validity is not obligatory for basic research, because the purpose in such research is not to generalize to practical situations but to make theoretical interpretations. A further advantage is that knowledge space – in combination with 4- and 5-generalizations – makes explicit the important point that a research investigation cannot operate in isolation, but needs help from earlier research in connection with inferences and their validities related to measurement and generalization of causal effects.

External validity in the Campbellian system and generalization validity in the COPS model are similar in that both concern whether a causal relationship holds over variation in causes, outcomes, persons, and settings. However, the involved causal relationship is defined at the indicator level in the system, but at the construct level in the model. Consequently, while external-validity generalizations in the system encompass also the special case where the study persons are randomly sampled from the study population (SCC, p. 84), this case corresponds to effect measurement instead of effect generalization in the model. For example, suppose the study sample in our therapy experiment has been randomly selected from the population of juvenile delinquents. The package effect of –30 holds over the observed and the unobserved adolescents (within the limits of sampling errors) in this population, and this case is therefore considered effect generalization in the system (SCC, p. 91). In contrast, this package effect is defined in terms of the actual C-, O-, P-, and S-construct classes in the model, and the case corresponds therefore here to effect measurement. It follows that the random sampling will strengthen external validity in the system, but will strengthen construct validity concerning persons and thereby strengthen causal validity in the model.

The random-sampling case above is considered by SCC (p. 91) a kind of ideal for external validity:

We have not put much emphasis on random sampling for external validity, primarily it is so rarely feasible in experiments. When it is feasible, however, we strongly recommend it, for just as random assignment simplifies internal validity inferences, so random sampling simplifies external validity inferences (assuming little or no attrition, as with random assignment).

However, the term “variation” in their external-validity definition (p. 38) should be interpreted as a systematic variation, while the difference between the two groups is due to random errors only. Consequently, since there is no real group difference here, their ideal random-sampling case can hardly be regarded as covered by their definition of external validity. A similar argument holds for random sampling of causes, outcomes, and settings.

A causal inference in ordinary research will include constructs, whereas a Campbellian internal-validity causal inference lacks constructs. This Campbellian inference has accordingly been criticized in the methodological literature for being unrepresentative of causal research. For example, as pointed out by Reichardt (2008), one consequence of this disadvantage is that the internal-validity concept is not applicable to the most common causal inferences that are drawn in research. Since causal construct-included inferences do not qualify as internal-validity inferences, they can neither be considered internally valid nor internally invalid. Such an inference could be wrong because of a threat to internal validity, but the inference itself cannot be characterized as internally invalid. A similar objection holds concerning a Campbellian external-validity generalization. Because such a generalization lacks constructs, whereas an ordinary causal generalization includes constructs, the external-validity concept is not applicable to the most common causal generalizations. A construct-included generalization can neither be considered externally valid nor externally invalid. Since internal-validity inferences and external-validity generalizations are unrepresentative of common causal inferences and causal generalizations, respectively, an internal-validity inference should be termed a *quasi-causal inference*, and an external-validity generalization should be termed a *quasi-*

*generalization*. Internal validity and external validity will then be properties of such quasi-causal inferences and quasi-generalizations, respectively. Finally, in the COPS model, causal inferences and generalizations include constructs, and they are therefore representative of common causal inferences and generalizations, respectively.

As mentioned, the COPS model is meant to apply not only to experimental and quasi-experimental studies, but also to non-experimental research. Suppose, for example, the research problem in some applied study is whether mothers' empathic communication with their children has a causal impact on the children's later social adjustment, using some non-experimental approach. According to the COPS model, the causal factor (mothers' empathic ...), the outcome (children's later ...), the children, the setting, and the control variables should be defined at the construct level, and the sampling particulars should be representative of their respective construct classes.

Cronbach (1982) has not only criticized the Campbellian system, but has also developed an alternative system. It is useful to compare his system with the COPS model. A short (and admittedly superficial) comparison is as follows: (1) His system is limited to applied research, especially evaluation research, and basic research is therefore not accounted for, whereas the model applies to both kinds of research. (2) The system includes three basic parts, namely *utos*, *UTOS*, and *\*UTOS*, where *u* and *U* stands for units of observation (e.g., persons), *t* and *T* for treatment, *o* and *O* for observation/outcome, and *s* and *S* for setting. Moreover, *utos* stands for sampling particulars in the research study, *UTOS* for the corresponding populations or construct classes which represent their respective sampling particulars, and *\*UTOS* for unstudied populations or construct classes of interest which differ from *UTOS* with respect to units, treatment, outcome, or setting. (3) *Internal validity* is defined in the system as the validity of the generalization from *utos* to *UTOS*. This validity can therefore be interpreted as somewhat parallel to causal validity in the model. (4) *External validity* is defined in the system as the validity of the generalization from *UTOS* to *\*UTOS*. This validity is therefore parallel to generalization validity in the model concerning unstudied variation. Since generalization validity applies also to studied variation, generalization validity is broader than Cronbach's external validity. (5) His external validity is considered more important than his internal validity. A corresponding priority between causal and generalization validity in the model is meaningless, because the latter validity depends on the former in applied research, and because generalization validity is not obligatory for basic research.

The preceding account has focused on validity. An inference should not only be valid, but also be relevant, and it can be considered relevant to the extent it represents needed knowledge. Such a requirement of relevance applies to all kinds of research, and therefore also to the Campbellian system and the COPS model. Satisfactory relevance is supposed in the account of the COPS model above. It should be added that while the validity concept can be divided into different validity categories, each supplied with methodological criteria for evaluating the category of validity, corresponding categories and criteria for relevance will be difficult to specify.

The account in this paper – as well as the criticism of the Campbellian validity system in the literature referred to above – lead to the conclusion that the system has several weaknesses and should be revised in line with the COPS model. That is, internal and external validity in the system should be replaced by causal validity and generalization validity in the model, and the system should be revised so it applies also to basic research.

## Disclosure Statement

No potential conflict of interest was reported by the author(s).

## References

Campbell, D. T. (1957). Factors relevant to the validity of experiments in social settings. *Psychological Bulletin*, 54(4), 297–312. <https://doi.org/10.1037/h0040950>

- Campbell, D. T. (1986). Relabeling internal and external validity for applied social scientists. In W. M. K. Trochim (Ed.), *Advances in quasi-experimental design and analysis* (pp. 67–77). Jossey-Bass.
- Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Rand-McNally.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for fields settings*. Houghton Mifflin Company.
- Cronbach, L. J. (1982). *Designing evaluations of educational and social programs*. Jossey-Bass.
- Guba, E., & Lincoln, Y. (1982). *Effective evaluation*. Jossey-Bass.
- Harlow, H. F. (1958). The nature of love. *American Psychologist*, 13(12), 673–685. <https://doi.org/10.1037/h0047884>
- Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Sage.
- Judd, C. M., & Kenny, D. A. (1981). *Estimating the effects of social interventions*. Cambridge University Press.
- Kenny, D. A. (1979). *Correlation and causality*. Wiley.
- Kruglanski, A. W., & Kroy, M. (1976). Outcome validity in experimental research: A re-conceptualization. *Journal of Representative Research in Social Psychology*, 7, 168–178.
- Lund, T. (2005). A metamodel of central inferences in empirical research. *Scandinavian Journal of Educational Research*, 49(4), 385–398. <https://doi.org/10.1080/00313830500202918>
- Lund, T. (2010). Causal inferences in the Campbellian validity system. *Scandinavian Journal of Educational Research*, 54(3), 205–220. <https://doi.org/10.1080/00313831003764495>
- Lund, T. (2013). Kinds of generalizations in educational and psychological research. *Scandinavian Journal of Educational Research*, 57(4), 445–456. <https://doi.org/10.1080/00313831.2012.657924>
- Mark, M. M. (1986). Validity typologies and the logic and practice of quasi-experimentation. In W. M. K. Trochim (Ed.), *Advances in quasi-experimental design and analysis* (pp. 47–66). Jossey-Bass.
- Mook, D. G. (1983). In defence of external invalidity. *American Psychologist*, 38(4), 379–387. <https://doi.org/10.1037/0003-066X.38.4.379>
- Reichardt, C. S. (2002). Review of Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and Quasi-experimental designs for generalized causal inference*. Houghton Mifflin. *Social Science Review*, 76, 510–514.
- Reichardt, C. S. (2006). The principle of parallelism in the design of studies to estimate treatment effects. *Psychological Methods*, 11(1), 1–18. <https://doi.org/10.1037/1082-989X.11.1.1>
- Reichardt, C. S. (2008, November). *An alternative to the Campbellian conceptualization of validity* [Paper presentation]. Evaluation 2008 Conference, Denver, CO, United States.
- Shadish, W. R. (2010). Campbell and Rubin: A primer and comparison of their approaches to causal inference in field settings. *Psychological Methods*, 15(1), 3–17. <https://doi.org/10.1037/a0015916>
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.
- Shadish, W. R., Cook, T. D., & Leviton, L. C. (1991). *Foundations of program evaluation: Theories of practice*. Sage.
- Smith, M. L., & Glass, G. V. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist*, 32(9), 752–760. <https://doi.org/10.1037/0003-066X.32.9.752>