Some Observations on Study Design

Author(s): Samuel A. Stouffer

Source: *American Journal of Sociology*, Jan., 1950, Vol. 55, No. 4 (Jan., 1950), pp. 355-361

Published by: The University of Chicago Press

Stable URL: https://www.jstor.org/stable/2772296

SOME OBSERVATIONS ON STUDY DESIGN

SAMUEL A. STOUFFER

ABSTRACT

Quick plausible "answers" in sociology and social psychology are rewarded in our culture; tedious, modest experimental design is not in demand, and hence our discipline is not cumulative. In study design the ideal model is that of a controlled experiment, even if only a fraction of it. Since full experimental design is very expensive and not always possible, those problems should be selected whose answers are worth the cost. This requires both theory which leads to operational deductions and preliminary fumbling research, whose intrinsic wastefulness can be reduced, if the number of variables is kept down to manageable limits and if such variables as are used are unidimensional.

As a youth I read a series of vigorous essays in the *Century Magazine* by its editor, the late Glenn Frank. His theme was that the natural sciences had remade the face of the earth; now had arrived the age of the social sciences. The same techniques which had worked their miracles in physics, chemistry, and biology should, in competent hands, achieve equally dazzling miracles in economics, political science, and sociology. That was a long time ago. The disconcerting fact is that people are writing essays just like that today. Of course, the last two decades have seen considerable progress in social science—in theory, in technique, and in the accumulation of data. It is true that the number of practitioners is pitifully few; only a few hundred research studies are reported annually in sociology, for example, as compared with more than twenty thousand studies summarized annually in *Biological Abstracts*. But the bright promise of the period when Frank was writing has not been fulfilled.

Two of the most common reasons alleged for slow progress are cogent, indeed.

The data of social science are awfully complex, it is said. And they involve values which sometimes put a strain on the objectivity of the investigator even when they do not incur resistance from the vested interests of our society. However, an important part of the trouble has very little to do with the subject matter of social science as such but, rather, is a product of our own bad work habits. That is why this paper on the subject of study design may be relevant. So much has been spoken and written on this topic that I make no pretense to originality. But in the course of a little experience, especially in an effort during the war to apply social psychology to military problems, and in an undertaking to nurture a new program of research in my university, I have encountered some frustrations which perhaps can be examined with profit.

A basic problem—perhaps *the* basic problem—lies deeply imbedded in the thoughtways of our culture. This is the implicit assumption that anybody with a little common sense and a few facts can come up at once with the correct answer on any subject. Thus the newspaper editor or columnist, faced with a column of empty space to fill with readable English in an hour, can speak with finality and authority on any social topic, however complex. He might not attempt to diagnose what is wrong with his sick cat; he would call a veterinarian. But he knows precisely what is wrong with any social institution and the remedies.

In a society which rewards quick and confident answers and does not worry about how the answers are arrived at, the social scientist is hardly to be blamed if he conforms to the norms. Hence, much social science is merely rather dull and obscure journalism; a few data and a lot of "interpretation." The fact that the so-called "interpretation" bears little or no relation to the data is often obscured by academic jargon. If the stuff is hard to read, it has a chance of being acclaimed as profound. The rewards are for the answers, however tediously expressed, and not for rigorously marshaled evidence.

In the army no one would think of adopting a new type of weapon without trying it out exhaustively on the firing range. But a new idea about handling personnel fared very differently. The last thing anybody ever thought about was trying out the idea experimentally. I recall several times when we had schemes for running an experimental tryout of an idea in the sociopsychological field. Usually one of two things would happen: the idea would be rejected as stupid without a tryout (it may have been stupid, too) or it would be seized on and applied generally and at once. When the provost marshal wanted us to look into the very low morale of the MP's, our attitude surveys suggested that there was room for very much better selectivity in job assignment. There were routine jobs like guarding prisoners which could be given to the duller MP's, and there were a good many jobs calling for intelligence, discretion, and skill in public relations. We thought that the smarter men might be assigned to these jobs and that the prestige of these jobs would be raised further if a sprinkling of returned veterans with plenty of ribbons and no current assignment could be included among them. We proposed a trial program of a reassignment system in a dozen MP outfits for the purpose of comparing the resulting morale with that in a dozen matched outfits which were left untouched. Did we get anywhere? No. Instead, several of our ideas were put into effect immediately throughout the army without any prior testing at all.

The army cannot be blamed for behavior like that. In social relations it is not the habit in our culture to demand evidence for an idea; plausibility is enough.

To alter the folkways, social science itself must take the initiative. We must be clear in our own minds what proof consists of, and we must, if possible, provide dramatic examples of the advantages of relying on something more than plausibility. And the heart of our problem lies in study design *in advance*, such that the evidence is not capable of a dozen alternative interpretations.

Basically, I think it is essential that we always keep in mind the model of a controlled experiment, even if in practice we may have to deviate from an ideal model. Take the simple accompanying diagram.

	Before	After	After − Before
Experimental group	x_1	x_2	$d = x_2 - x_1$
Control group	x'_1	x'_2	$d' = x'_2 - x'_1$

The test of whether a difference d is attributable to what we think it is attributable to is whether d is significantly larger than d'.

We used this model over and over again during the war to measure the effectiveness of orientation films in changing soldiers' attitudes. These experiences are described in Volume III of our *Studies in Social Psychology in World War II*.[1]

One of the troubles with using this careful design was that the effectiveness of a single film when thus measured turned out to be so slight. If, instead of using the complete experimental design, we simply took an unselected sample of men and compared the attitudes of those who said they had seen a film with those who said they had not, we got much more impressive differences. This was more rewarding to us, too, for the management wanted to believe the films were powerful medicine. The gimmick was the selective fallibility of memory. Men who correctly remembered seeing the films were likely to be those most sensitized to their message. Men who were bored or indifferent may have actually seen them but slept through them or just forgot.

Most of the time we are not able or not patient enough to design studies containing all four cells as in the diagram above. Some-

[1] Carl I. Hovland, Arthur A. Lumsdaine, and Fred D. Sheffield, *Experiments in Mass Communication* (Princeton: Princeton University Press, 1949).

times we have only the top two cells, as in the accompanying diagram. In this situation

x_1	x_2

$d = x_2 - x_1$

we have two observations of the same individuals or groups taken at different times. This is often a very useful design. In the army, for example, we would take a group of recruits, ascertain their attitudes, and restudy the same men later. From this we could tell whose attitudes changed and in what direction (it was almost always for the worse, which did not endear us to the army!). But exactly what factors in the early training period were most responsible for deterioration of attitudes could only be inferred indirectly.

The panel study is usually more informative than a more frequent design, which might be pictured thus:

x_1	
	x_2'

Here at one point in time we have one sample, and at a later point in time we have another sample. We observe that our measure, say, the mean, is greater for the recent sample than for the earlier one. But we are precluded from observing which men or what type of men shifted. Moreover, there is always the disturbing possibility that the populations in our two samples were initially different; hence the differences might not be attributable to conditions taking place in the time interval between the two observations. Thus we would study a group of soldiers in the United States and later ask the same questions of a group of soldiers overseas. Having matched the two groups of men carefully by branch of service, length of time in the army, rank, etc., we hoped that the results of the study would approximate what would be found if the same men could have been studied twice. But this could be no more than a hope. Some important factors could not be adequately controlled, for example, physical conditions. Men who went overseas were initially in better shape on the average than men who had been kept behind; but, if the follow-up study was in the tropics, there was a chance that unfavorable climate already had begun to take its toll. And so it went. How much men overseas changed called for a panel study as a minimum if we were to have much confidence in the findings.

A very common attempt to get the results of a controlled experiment without paying the price is with the design that might be as shown in the accompanying diagram. This

x_2
x_2'

is usually what we get with correlation analysis. We have two or more groups of men whom we study at the same point in time. Thus we have men in the infantry and men in the air corps and compare their attitudes. How much of the difference between x_2' and x_2 we can attribute to experience in a given branch of service and how much is a function of attributes of the men selected for each branch we cannot know assuredly. True, we can try to rule out various possibilities by matching; we can compare men from the two branches with the same age and education, for example. But there is all too often a wide-open gate through which other uncontrolled variables can march.

Sometimes, believe it or not, we have only one cell:

x_2

When this happens, we do not know much of anything. But we can still fill pages of social science journals with "brilliant analysis" if we use plausible conjecture in supplying missing cells from our imagination. Thus we may find that the adolescent today has wild ideas and conclude that society is going to the dogs. We fill in the dotted cell representing our own yesterdays with hypothetical data, where x_1 represents us and x_2 our off-

x_1	x_2

spring. The tragicomic part is that most of the public, including, I fear, many social scientists, are so acculturated that they ask for no better data.

I do not intend to disparage all research not conforming to the canons of the controlled experiment. I think that we will see more of full experimental design in sociology and social psychology in the future than in the past. But I am well aware of the practical difficulties of its execution, and I know that there are numberless important situations in which it is not feasible at all. What I am arguing for is awareness of the limitations of a design in which crucial cells are missing.

Sometimes by forethought and patchwork we can get approximations which are useful if we are careful to avoid overinterpretation. Let me cite an example:

In Europe during the war the army tested the idea of putting an entire platoon of Negro soldiers into a white infantry outfit. This was done in several companies. The Negroes fought beside white soldiers. After several months we were asked to find out what the white troops thought about the innovation. We found that only 7 per cent of the white soldiers in companies with Negro platoons said that they disliked the idea very much, whereas 62 per cent of the white soldiers in divisions without Negro troops said they would dislike the idea very much if it were tried in their outfits. We have:

	Before	After
Experimental		7%
Control		62%

Now, were these white soldiers who fought beside Negroes men who were naturally more favorable to Negroes than the cross-section of white infantrymen? We did not think so, since, for example, they contained about the same proportion of southerners. The point was of some importance, however, if we were to make the inference that actual experience with Negroes reduced hostility from 62 to 7 per cent. As a second-best substitute, we asked the white soldiers in companies with Negro platoons if they could recall how they felt when the innovation was first proposed. It happens that 67 per cent said they were initially opposed to the idea. Thus we could tentatively fill in a missing cell and conclude that, under the conditions obtaining, there probably had been a marked change in attitude.

Even if this had been a perfectly controlled experiment, there was still plenty of chance to draw erroneous inferences. The conclusions apply only to situations closely approximating those of the study. It happens, for example, that the Negroes involved were men who volunteered to leave rear-area jobs for combat duty. If other Negroes had been involved, the situation might have been different. Moreover, they had white officers. One army colonel who saw this study and whom I expected to ridicule it because he usually opposed innovations, surprised me by offering congratulations. "This proves," he said, "what I have been arguing in all my thirty years in the army—that niggers will do all right if you give 'em white officers!" Moreover, the study applied only to combat experience. Other studies would be needed to justify extending the findings to noncombat or garrison duty. In other words, one lone study, however well designed, can be a very dangerous thing if it is exploited beyond its immediate implications.

Now experiments take time and money, and there is no use denying that we in social science cannot be as prodigal with the replications as the biologist who can run a hundred experiments simultaneously by growing plants in all kinds of soils and conditions. The relative ease of experimentation in much—not all—of natural science goes far to account for the difference in quality of proof demanded by physical and biological sciences, on the one hand, and social scientists, on the other.

Though we cannot always design neat experiments when we want to, we can at least keep the experimental model in front of our eyes and behave cautiously when we fill in missing cells with dotted lines. But there is a further and even more important operation we can perform in the interest of economy. That lies in our choice of the initial problem.

Professor W. F. Ogburn always told his students to apply to a reported research conclusion the test, "How do you know it?" To this wise advice I should like to add a further question: "What of it?" I suspect that if before designing a study we asked ourselves, more conscientiously than we do, whether or not the study really is important, we would economize our energies for the few studies which are worth the expense and trouble of the kind of design I have been discussing.

Can anything be said about guides for selecting problems? I certainly think so. That is where theory comes in and where we social scientists have gone woefully astray.

Theory has not often been designed with research operations in mind. Theory as we have it in social science serves indispensably as a very broad frame of reference or general orientation. Thus modern theories of culture tell us that it is usually more profitable to focus on the learning process and the content of what is learned rather than on innate or hereditary traits. But they do not provide us with sets of interrelated propositions which can be put in the form: If x_1, given x_2 and x_3, then there is strong probability that we get x_4. Most of our propositions of that form, sometimes called "theory," are likely to be *ad hoc* common-sense observations which are not deducible from more general considerations and which are of the same quality as the observation, "If you stick your hand in a fire and hold it there, you will get burned."

Now in view of the tremendous cost in time and money of the ideal kind of strict empirical research operations, it is obvious that we cannot afford the luxury of conducting them as isolated fact-finding enterprises. Each should seek to be some sort of *experimentum crucis*, and, with rare exceptions, that will only happen if we see its place *beforehand* in a more general scheme of things. Especially, we need to look for situations where two equally plausible hypotheses deducible from more general theory lead to the expectation of different consequences. Then, if our evidence supports one and knocks out the other, we have accomplished something.

The best work of this sort in our field is probably being done today in laboratory studies of learning and of perception. I do not know of very good sociological examples. Yet in sociology experiments are possible. One of the most exciting, for example, was that initiated long before the war by Shaw and McKay to see whether co-operative effort by adult role models within a delinquent neighborhood would reduce juvenile delinquency. So many variables are involved in a single study like that that it is not easy to determine which were crucial. But there was theory behind the study, and the experimental design provided for controlling at least some variables.

It may be that in sociology we will need much more thinking and many more descriptive studies involving random ratlike movements on the part of the researcher before we can even begin to state our problems so that they are in decent shape for fitting into an ideal design. However, I think that we can reduce to some extent the waste motion of the exploratory period if we try to act as if we have some a priori ideas and keep our eyes on the possible relevance of data to these ideas. This is easier said than done. So many interesting rabbit tracks are likely to be uncovered in the exploratory stages of research that one is tempted to chase rabbits all over the woods and forget what his initial quarry was.

Exploratory research is of necessity fumbling, but I think that the waste motion can be reduced by the self-denying ordinance of deliberately limiting ourselves to a few variables at a time. Recently two of my colleagues and myself have been doing a little exploratory work on a problem in the general area of social mobility. We started by tabulating some school records of fifty boys in the ninth grade of one junior high school and then having members of our seminar conduct three or four interviews with each boy and his parents. We had all the interviews written up in detail, and we had enough data to fill a book—with rather interesting reading, too. But it was a very wasteful process because there were just too many intriguing ideas. We took a couple of ideas which were deducible from current general theory and tried to make some simple fourfold tables. It was obvious that, with a dozen variables uncontrolled, such tables meant little or nothing. But that led us to a second step. Now we are trying to collect school records and a short questionnaire on two thousand boys. We will not interview all these boys and their parents in detail. But, with two thousand cases to start with, we hope to take a variable in which we are interested and find fifty boys who are plus on it and fifty who are minus, yet who are approximately alike on a lot of other things. A table based on such matched comparisons should be relatively unambiguous. We can take off from there and interview those selected cases intensively to push further our exploration of the nexus between theory and observation. This, we think, will be economical, though still exploratory. Experimental manipulation is far in the future in our problem, but we do hope we can conclude the first stage with a statement of some hypotheses susceptible to experimental verification.

I am not in the least deprecating exploratory work. But I do think that some orderliness is indicated even in the bright dawn of a youthful enterprise.

One reason why we are not more orderly in our exploratory work is that all too often what is missing is a sharp definition of a given variable, such that, if we wanted to take a number of cases and even throw them into a simple fourfold table, we could.

Suppose we are studying a problem in which one of the variables we are looking for is overprotection or overindulgence of a child by his mother. We have a number of case histories or questionnaires. Now how do we know whether we are sorting them according to this variable or not? The first step, it would seem, is to have some way of knowing whether we are sorting them along any single continuum, applying the same criteria to each case. But to know this we need to have built into the study the ingredients of a scale. Unless we have some such ingredients in our data, we are defeated from the start. This is why I think the new interest social scientists are taking in scaling techniques is so crucially important to progress. In particular, the latent-structure theory developed by Paul F. Lazarsfeld, which derives Louis Guttman's scale as an important special case, is likely to be exceedingly useful, for it offers criteria by which we can make a small amount of information go a long way in telling us the logical structure of a supposed variable we are eager to identify. The details of Guttman's and Lazarsfeld's work[2] are likely to promote a good deal of attack and controversy. Our hope is that this will stimulate others to think such problems out still better and thus make their work obsolete as rapidly as possible.

Trying to conduct a social science investigation without good criteria for knowing whether a particular variable may be treated as a single dimension is like trying to fly without a motor in the plane. Students of the history of invention point out that one reason why the airplane, whose properties had been pretty well thought out by Leonardo da Vinci, was so late in development was the unavailability of a lightweight power plant, which had to await the

[2] Samuel A. Stouffer, Louis Guttman, Edward A. Suchman, Paul F. Lazarsfeld, Shirley A. Star, and John A. Clausen, *Measurement and Prediction* (Princeton: Princeton University Press, 1949).

invention of the internal combustion motor. We are learning more and more how to make our light-weight motors in social science, and that augurs well for the future. But much work is ahead of us. In particular, we desperately need better projective techniques and better ways of getting respondents to reveal attitudes which are too emotionally charged to be accessible to direct questioning. Schemes like the latent-structure theory of Lazarsfeld should speed up the process of developing such tests.

I have tried to set forth the model of the controlled experiment as an ideal to keep in the forefront of our minds even when by necessity some cells are missing from our design. I have also tried to suggest that more economy and orderliness are made possible, even in designing the exploratory stages of a piece of research—by using theory in advance to help us decide whether a particular inquiry would be important if we made it; by narrowing down the number of variables; and by making sure that we can classify our data along a particular continuum, even if only provisionally. And a central, brooding hope is that we will have the modesty to recognize the difference between a promising idea and proof.

Oh, how we need that modesty! The public expects us to deal with great problems like international peace, full employment, maximization of industrial efficiency. As pundits we can pronounce on such matters; as citizens we have a duty to be concerned with them; but as social scientists our greatest achievement now will be to provide a few small dramatic examples that hypotheses in our field can be stated operationally and tested crucially. And we will not accomplish that by spending most of our time writing or reading papers like this one. We will accomplish it best by rolling up our sleeves and working at the intricacies of design of studies which, though scientifically strategic, seem to laymen trivial compared with the global concerns of the atomic age. Thereby, and only thereby, I believe, can we some day have the thrilling sense of having contributed to the structure of a social science which is cumulative.

HARVARD UNIVERSITY