

CASE STUDY

The Hawthorne Effect

In the 1920s the Western Electric Company had a factory at Cicero, Illinois, just outside Chicago. Between 1924 and 1932, a series of field studies and experiments were carried out on workers at the plant. The research was funded by the National Research Council of the National Academy of Sciences and undertaken on behalf of General Electric, the largest manufacturer of light bulbs in the United States. The aim was to assess the influence of lighting on workforce productivity, and therefore to determine the level of illumination which produced the best work performance.

When the researchers discovered that productivity almost always increased after a change in illumination, regardless of whether the light was increased or decreased, they performed a second set of experiments, supervised by Harvard University professors Elton Mayo, Fritz Roethlisberger and William J. Dickson, in order to find out why their measurement methods were unreliable.

They introduced other changes into the work environment and studied the responses to them by a sample of five young women. Once again, whatever change was made to the conditions, the women's productivity almost invariably improved.

These studies engendered the expression 'the Hawthorne effect': an important advance in industrial and organizational psychology and in organizational behavior, whose general meaning was that when people are being observed by researchers, they temporarily change their behavior or performance. The definition has since been extended to state that people change their behaviors and performances in response to *any* increase in the attention paid to them.

There is another, even more insidious, aspect of reliability. It concerns cases in which the instrument is intrinsically unsuited to the kind of research being undertaken. In his field studies on the linguistic features of African American Vernacular English (AAVE), now called Ebonics, William Labov noted that the poor language ability displayed by black boys in interviews was a highly unsatisfactory indicator of their real linguistic skills. This was because the boys were obliged to express themselves in a communicative modality which they found unfamiliar, and which also provoked their hostility:

If one takes (the) interview as a measure of the verbal capacity of the child, it must be as his capacity to defend himself in a hostile and threatening situation. But unfortunately, thousands of such interviews are used as evidence of the child's total verbal capacity, or more simply his verballity: it is argued that this lack of 'verballity' explains his poor performance in school (1972: 61).

Suppose that we have made two or more observations (measurements) yielding the same result. For example: we have put the same question, at different times, to the same participant or informant, and we have always received the same reply.

Can we say that the finding is necessarily true (i.e. valid) because it is consistent and replicable? No, we cannot. The participant or informant may have lied on both occasions. Or the question may have been formulated wrongly, so that it failed to elicit the true information. Herein resides the entire paradox of reliability: the data may be erroneous but the method itself is reliable. Even worse, the instrument itself may be faulty but nonetheless perfectly reliable – as in the following example given by the quantitative methodologists Edward G. Carmines and Richard A. Zeller:

let us assume that a particular yardstick does not equal 36 inches; instead, the yardstick is 40 inches long. Thus, every time this yardstick is used to determine the height of a person (or object), it systematically underestimates height by 4 inches for every 36 inches . . . however, this error of 4 inches per yard will *not affect* the reliability of the yardstick since it does not lead to inconsistent results on repeated measurements . . . In short, this particular yardstick will provide a quite reliable but totally invalid indication of height (1979: 13).

This is why reliability, though certainly useful, is a reductive concept which must rest on other criteria, and among them (but not solely) validity.