Evaluation of intervention measures  
– a methodological approach

A. Karltn

Division of Industrial Ergonomics, Linköping University, Linköping, Sweden

Abstract

It is common for researchers to evaluate interventions by having the participants answer standardized questions before and after, and measuring the effect of the intervention as a statistical difference between the two. There is a risk that the results of such a calculation will be misleading due to the subject interpreting the response scales differently on each occasion. This risk has previously been observed by Golembiewski et al, who indicated that we are dealing with a psychometric problem set that is especially evident in interventions. The question is whether the individual’s own yardstick for assessing or valuing a certain circumstance/area is the same before and after the intervention, or whether the yardstick has changed due to the subject gaining new experience within the conceptual framework. Golembiewski calls this phenomenon beta change. The purpose of this paper is to illustrate empirically the occurrence of beta change and its possible consequences for the interpretation of the results, with the aid of a model devised by Terborg et al. The material is taken from a health-promoting intervention project in the mail distribution division of the Swedish Post Office.

Keywords: intervention research, measurement problems, change, psychometrics, questionnaire

1. Introduction

The Swedish Post Office has been the subject of large-scale restructuring since the mid 1990s. For the mail delivery division, this has meant changes in order to achieve better, more efficient work processes at a lower cost. The division’s management called in the Division of Industrial Ergonomics at Linköping University to collaborate in different phases with the aim of instituting efficient, health-promoting work processes. Improvement measures were proposed, constituting the foundation of a major project and eighteen months later solutions had been devised that were ready for implementation: a) new labelling and lighting for sorting racks, b) adjustment and optimisation of the sorting racks and ancillary space, c) training in better work techniques for sorting mail, d) manual for organising efficient indoor work at the local delivery offices. To begin with, this was done in the form of a pilot intervention study at a delivery office in Lidingö [4]. Implementation began in early October and continued until the first half of May 2004.

2. On the evaluation of interventions - a measurement problem

A common way of evaluating an intervention study is to divide it time-wise into three main phases: the first before the intervention begins, the second its actual implementation, and the third on a number of occasions after its completion [1]. Since an intervention is by its very nature a deliberate action to improve something, one of its common effects is to change, enhance, or colour the subjects’ understanding and opinions of the problem along the way. The
intervention will often bring about an ongoing learning process that at the same time gives rise to a measurement-related dilemma that can be expressed as follows [5]: “When scores on a test are observed to change, how can one tell whether it is the persons who have changed or the test?” A questionnaire is often used in evaluations and it is common for researchers to have the participants answer the same standardized questions on repeated occasions and measure the effect of the intervention as a statistical difference between the before and after measurements. We then run the risk of encountering a beta change as defined by Golembiewski [2]: “BETA CHANGE involves a variation in the level of some existential state, complicated by the fact that some intervals of the measurement continuum associated with a constant conceptual domain have been recalibrated.” The opposite pole to such a change in the subjective yardstick is when the respondent’s yardstick remains unchanged. This is what Golembiewski calls an alpha change [2]: “ALPHA CHANGE involves a variation in the level of some existential state, given a constantly calibrated measuring instrument related to a constant conceptual domain.

Westlander [6] has this to say: If we view the survey-based, applied intervention research as results delivered up until the present time, it is obvious that what Golembiewski defines as alpha change is the type of change that most researchers reckon with. …/ Beta change, on the other hand is, a ‘perceived’ problem that is discussed and undeniably causes some concern among some intervention researchers: should we have had more alternative answers, ‘stretched’ the response scale, had more steps on the scale, defined the limits in the direction of even more extreme in order to capture the respondents’ opinions?”

In light of this, it is of interest to seek answers to the following questions: Can we see any indications of beta change in the evaluation of the four measures taken in the pilot study at the delivery office in Lidingö? What are the consequences for the results as regards the reliability of results that show a difference between before and after measurements? Analysis at the group level, i.e. comparisons of differences in the group mean between before and after, appears to be a common solution in major intervention studies where the risk of dropout after the pre-intervention stage is as good as inevitable [6, 2, 7]. Terborg et al [3] state that analysis solely at group level can be misleading particularly if the purpose is to support an intervention practically. They have therefore put forward a method that analyses beta change at the level of the individual.

According to Terborg and his co-authors [3], the individual level is to be preferred to the group level for several reasons. However, it requires subjects to state their identity (name or something else) at the time of both before and after measurement. The first reason for analysis at the individual level is that relatively large changes in mean values for a few group members may lead the researcher to conclude that the intervention has had an effect at group level, while in reality it may only have been successful for some parts of the group. Second, it may be important, especially if several categories in an organisation are concerned, to distinguish those individuals for whom changes have occurred and in what direction. Third, analyses at the individual level allow the researcher to see whether the intervention has brought about different types of change in different individuals. Group level analyses mask such differentiations. A further advantage is that the method allows a small number of participants to be handled, something that is common in interventions. Finally, knowledge of changes at the individual level can give the intervention researcher an enhanced understanding of the effects of the interventions, at the same time as it provides a better basis for feedback of the results during and after the intervention process.

A beta change according to the model put forward by Terborg et al [3] is based on three measurements: PRE, POST, and THEN. These terms refer to one measurement before the intervention (PRE), and two made at the same time after the intervention - how the situation is perceived now (POST) and before the intervention (THEN). The retrospective THEN measurement is always made at the same time as the POST measurement and using the same category scale. The method is thus a way of showing whether each individual’s perceptions before the interventions are “remembered correctly” afterwards, which is a crucial factor as regards whether any conclusions can be drawn about the effect of an intervention expressed as the difference between before and after measurements.

A beta change is reflected in the degree of difference between the individual’s values in the PRE and THEN measurements. If the responses for PRE and THEN do not differ, this means, according to Terborg et al [3], that the difference between the response in the POST measurement and the response in the PRE measurement can be trusted; in other words no beta change has occurred. On the other hand, if there is a difference between the responses in the PRE and THEN measurements, this means that it is not possible to draw any conclusion solely on the basis of
the difference between only PRE and POST. POST may be coloured by the THEN recollection and the individual may have recalibrated his or her own subjective yardstick. If no beta change can be determined through comparison between PRE and THEN, we may permit ourselves to rely on the PRE-POST discrepancy and understand that we have encountered an alpha change according to Golembiewski’s definition. Terborg et al [3] also discuss reservations proposed by other authors with regard to THEN measurements such as difficulties in remembering correctly, the possible effect of a retrospective THEN measurement on the POST measurement and a hypothetical risk that the respondents’ answers might be influenced by social pressure to perform better in the THEN measurement. The point of their model, however, is that whether the THEN recollection differs from the PRE perception or not, it has relevance for the POST valuation, which is always related to the reference that the interview subject has.

3. Method and material

A step-wise implementation plan for introducing and evaluating the various improvement measures was drawn up and approved by the staff at the delivery office in Lidingö. Before each step the postmen were asked to answer a PRE question, and after about 14 days of working under the new conditions a THEN and a POST question. The questions were valuation questions and sensitive to frames of reference and courses of events – they could not be “correct” purely in the sense of right or wrong. Respondents answered the questions on a scale from one to five.1 Every questionnaire was given a serial number that was then used throughout for each questionnaire and postman. The requirement for investigating the possible occurrence of beta change was that every participant answered all three questions reported here for each of the four improvement measures in the measurement series. 15 of the 35 postmen met this requirement. The questionnaire data from the PRE, POST, and THEN measurements for fifteen subjects were analysed at the individual level. After the fifth and final part questionnaire, a series of group interviews were conducted and analyzed describing the work team’s perception on the basis of the improvement measures, viewed as a whole.

4. Results

Can we see any indications of beta change in the evaluation of the four measures in the pilot delivery office? What are the consequences for the results as regards the reliability of results that show a difference between before and after measurements? Before answering these questions, let us begin by answering the question whether the questionnaire responses show any group tendency as regards a change in perception between before the measure was taken (PRE) and after it had been implemented (POST), and a change between the before measurement (PRE) and the recollection afterwards (THEN). The response scale in the questionnaire is from 1 to 5, where 1 represents the most positive response and 5 the most negative.

In a comparison between the group means for the PRE and POST measurements, differences were noted in all four measurement series (table 1), but a significant difference between PRE and POST measurement was only obtained in respect of the action taken to improve the lighting.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Mean value PRE</th>
<th>Mean value POST</th>
<th>Discrepancy PRE-POST</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manual</td>
<td>2.93</td>
<td>2.73</td>
<td>+0.20</td>
</tr>
<tr>
<td>Labelling</td>
<td>2.67</td>
<td>2.13</td>
<td>+0.54</td>
</tr>
<tr>
<td>Lighting</td>
<td>3.20</td>
<td>1.73</td>
<td>+1.47*</td>
</tr>
<tr>
<td>Work techniques</td>
<td>2.60</td>
<td>2.87</td>
<td>-0.27</td>
</tr>
</tbody>
</table>

Wilcoxon signed-rank significance test. * p < .01

A difference in the mean values between PRE and THEN can be seen in the measurements for the manual and the labelling, while lighting and work techniques show no differences (table 3).

<table>
<thead>
<tr>
<th>Measure</th>
<th>Mean value PRE</th>
<th>Mean value THEN</th>
<th>Discrepancy PRE-THEN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manual</td>
<td>2.93</td>
<td>2.80</td>
<td>+0.13</td>
</tr>
<tr>
<td>Labelling</td>
<td>2.67</td>
<td>2.20</td>
<td>+0.47</td>
</tr>
<tr>
<td>Lighting</td>
<td>3.20</td>
<td>3.20</td>
<td>+/- 0</td>
</tr>
<tr>
<td>Work techniques</td>
<td>2.60</td>
<td>2.60</td>
<td>+/- 0</td>
</tr>
</tbody>
</table>

Wilcoxon signed-rank significance test.
We now come to the stage where we determine whether a beta change has occurred or not by first investigating whether the THEN recollection differs from PRE, i.e. how the fifteen subjects assessed the situation before the measure was implemented. We begin by showing the conformity between PRE perception and THEN recollection for each of the fifteen subjects.

Table 3. Degree of conformance between PRE perception and THEN recollection for the 15 subjects.

<table>
<thead>
<tr>
<th></th>
<th>More positive recollection</th>
<th>More negative recollection</th>
<th>Perception unchanged</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manual</td>
<td>5</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Labelling</td>
<td>6</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Lighting</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Work techniques</td>
<td>5</td>
<td>6</td>
<td>4</td>
</tr>
</tbody>
</table>

Table 3 shows that more subjects have changed their opinion than have not after the measures were implemented, compared to how they assessed the situation as it was before. Most of the 15 subjects display differences in the measurement values between PRE and THEN in all measurement series. As can be seen from table 3, THEN recollection is spread fairly evenly in both the positive and the negative directions relative to the PRE measurement. Approximately one third of the subjects did not change their opinion about the PRE situation. In 40 of the 60 THEN measurements, a different response was given compared to the PRE measurement. The analysis gave some indication that a beta change has taken place in most of the individuals in the study and that we must take this into consideration when we interpret the differences between the values in the PRE and POST measurements.

When the results were communicated to the respondents, they said that the only reasonable explanation for the difference between the PRE and THEN values that they could think of was that their frame of reference had been influenced by each improvement measure over the course of the project, and that this had caused changes in their personal yardstick between the PRE and THEN measurements. The differences in group mean value that were observed can thus be regarded as obviously masking and misleading.

The continued follow-up of the measures implemented at the delivery office opened possibilities to further understand the individual differences in the reactions according to the standardized questions. Different conceivable indications as to why the response patterns vary like this can be obtained from the work team interviews and observations made by the author during the course of implementing the measures.

5. Conclusions and discussion

We can see that a beta change has taken place in most of the fifteen subjects who took part in this particular study. It is evident that they have recalibrated their subjective yardsticks between the times the measurements were made. We should therefore draw the conclusion that a difference between PRE and POST results is not reliable since it does not reflect the change between the before and after measurements correctly. This brings to the fore the primary motive put forward by Terborg et al [3] for analysis at the individual level according to the above, i.e. that relatively large changes in group mean values based on a few individuals may lead the researcher to conclude that the intervention has had an unambiguous effect at group level, while in reality it may only have been successful for some parts of the group.

We cannot fully reject alpha change. Approximately one third of the individuals display an alpha change while two thirds are examples of beta change. As far as can be judged, different types of change have occurred in different individuals. In those individuals where beta change has been observed, the change was in different directions for different individuals. The results give us important information about how we can differentiate future measures, which Terborg [3] cites as a second and a third reason for analysis at the individual level. Group level analyses mask such differentiations.

We might ask ourselves whether it would not be more relevant to keep to a comparison between the THEN and POST values in order to obtain a correct picture of the effects of the intervention thus far and the thrust that future measures should be given. And why not instead word the questions themselves in terms of the perceived difference between the conditions before and after implementing the interventions?

First, this would make it impossible to measure the difference between different types of change (alpha or beta) and in which individuals they occur. It would make it difficult to “separate apples from pears” and would thus not provide such a good basis for adapting
future measures. Second, we need to ask ourselves what we wish to measure in each specific case. The very purpose of many interventions and perhaps organisational interventions in particular, is to affect individuals’ frames of reference and by extension their behavioural patterns. In other words, it is very valuable to be able to determine whether a beta change has occurred or not.

In the light of the follow-up interviews, it is quite clear that many individuals have changed their opinion, for example with regard to what good work lighting is once they had worked with the new lighting fixtures for a time. Without the THEN measurement, the measurement values for a subject whose responses were PRE=1, THEN=4, and POST=1, would be interpreted as indicating that no change has taken place between PRE and POST. When we have access to the THEN value we can see that something has happened to the individual’s frame of reference during the implementation of the change and the individual’s estimation then has a completely different implication. Different individuals were affected in different ways by the four improvement measures depending on circumstances that we are – thanks to the interview material obtained – now better equipped to deal with in practice. This approach also made it possible to handle and make an exhaustive analysis of a relatively small number of subjects, as is often the case with interventions and which advocates of the method put forward as a further advantage.

Quite naturally, the model propounded by Terborg et al [3] does not go unchallenged. An objection in principle is put forward by Van de Vliert et al [8], referring to the fact that an external validation criterion needs to be applied to the results of the questionnaire. No single method is perfect or the one true faith. One thing, however, is important: knowing what one wishes to determine from a measurement and what we can believe from the results. This minimises the risk of obtaining results from something other than what we intended to measure. The validity of this study’s results can be tested in subsequent individual interviews.

Like Terborg et al [3], the author would also like to claim that this methodological approach can contribute to a fuller, deeper understanding of the effects of an intervention. The results generate a platform from which to act when communicating the results during and after the intervention process. At the same time, it is a challenge for the intervention researcher to decide how feedback of the results can be reflected in this deeper understanding of the effects of the intervention.

Acknowledgements

The author wants to thank Professor Gunnela Westlander for her methodological support and comments on earlier versions of this paper.

References