Evaluating training courses: an exercise in social desirability?

This item was submitted to Loughborough University’s Institutional Repository by the/an author.


Additional Information:

- This article was published in the journal, Journal of European Industrial Training [© Emerald] and is also available at: http://www.emeraldinsight.com/info/journals/jeit/jeit.jsp.

Metadata Record: [https://dspace.lboro.ac.uk/2134/2555](https://dspace.lboro.ac.uk/2134/2555)

Publisher: © Emerald

Please cite the published version.
This item was submitted to Loughborough’s Institutional Repository by the author and is made available under the following Creative Commons Licence conditions.

For the full text of this licence, please go to:
http://creativecommons.org/licenses/by-nc-nd/2.5/
EVALUATING TRAINING COURSES: AN EXERCISE IN SOCIAL DESIRABILITY?

Jenny A. Darby

Department of Social Sciences, Loughborough University, Loughborough, Leicestershire. LE11 3TU

Abstract:

Purpose: This study examines factors, which influence responses on open ended evaluations of training courses.

Method: Course participants completed open ended evaluation forms about their experience on a course. These consisted of 377 senior teachers attending a training programme dealing with child abuse. The course was repeated 17 times. The second training programme concerned teaching skills. This was attended by 231 postgraduates. The course was repeated 25 times.

Findings: Responses on open ended evaluation forms tended to be favourable with reference to ‘human related factors’ and unfavourable when referring to ‘hygiene factors’.

Implications: It is suggested the way people complete evaluation forms is partly a reflection of their desire to see themselves as acting in a socially desirable manner. Interpretations made from such forms about the effectiveness or merits of any course should take this into account.

Key words: evaluation, social influence, student, training, open-ended.

Introduction

Evaluation is commonly regarded as an essential aspect of any training course. MacIntyre and Carr (2000) point out the majority of course organisers do evaluate, and this is frequently an attempt to respond to external pressure to demonstrate managerial efficiency, quality and accountability in training. Warr, Bird and Rackham (1971) summarise the benefits of evaluating in terms of justifying the training itself, enhancing the value of individual courses and increasing the future effectiveness of training. They also draw attention to the need for evaluating the transfer of learning to performance in the work place. Rae (2002) however accepts the reality that, for many trainers pencil and paper evaluations immediately courses are completed is the best that can be achieved. One style of this latter type of evaluation is the open ended questionnaire. Typically favourable or unfavourable trigger words are provided as a stimulus and participants are expected to provide their first impressions of the course. These trigger words include phrases such as ‘the best/worst’ thing, ‘what I liked most/least’, ‘what helped/hindered’ about the course.

This style of evaluation questionnaire is the subject of the present study.

An awareness of acting in a socially desirable manner

Our reactions to situations are frequently influenced by a desire to be seen by others in a ‘good light’. This really means to be seen as being socially acceptable or ‘fitting in’. Those involved with Human Resource Management will be very familiar with this. The tendency to give socially desirable responses has been noted in many research methodology texts as a major problem, for example when interviewing someone (eg. Bryman 2001, p 112, Coleman and Briggs 2002, p 143, Fowler 2002, Hayes 2000 Memon and Bull 2000 p221, Shaughnesssey et al 2000 and Sommer and Sommer 1997, p 119). They all point out that interviewees attempt to find socially desirable ways to respond to questions to improve their image. Individuals seek to show themselves in a ‘good light’ hence it important questions are carefully worded. They will make statements they think comply with the norms of their group, or even the expectations of the interviewer.

A failure to link social desirability with training evaluations

The same texts however, fail to link the idea of reacting in a socially desirable manner with course evaluations. For example when Bryman (2001, p 143) and Sommer and Sommer (1997, p 128-130), provide a background explanation of open ended questionnaires no mention is made of the possibility of responses being influenced by social desirability. Erickson and Kaplan (2000) go even further and mention the virtues of open ended survey questions which they say ‘often elicit more honest responses’.
Trochim (2002) suggest open ended questionnaires can ‘provide a rich description of respondents reality’. Nowhere in either of these last two studies do they mention the respondents may be responding in a socially desirable manner. Indeed there is a long history of researchers actually claiming questionnaires are not subject to pressures to respond in a socially desirable manner. Diamond (1972), suggested reacting in a socially desirable manner only occurs when the person completing the form can be identified. With questionnaires the assumption seems to be summarised by Sudman and Bradburn (1982) that individuals will be honest if the responses are anonymous and confidential. They suggest considerations of social desirability and self preservation can, by these means, be eliminated. This notion is supported by more recent researchers, such as Strack et al (1990), who argue that by making response forms anonymous the likelihood of social desirable responses occurring is greatly reduced.

Favourable and unfavourable reactions in situations
When evaluating, both favourable and unfavourable experiences in courses are often mentioned by participants. There is evidence in the literature, taken from a variety of very different areas of research, which suggests individuals react differently to achieve socially desirable reactions when responding in a favourable or unfavourable manner. This is of concern in this study. Monson, Tanke and Lund (1980) note that people do not like to be negative about others, even if it is not face to face. This is particularly pertinent with regard to this present study where evaluation forms are completed anonymously. There is further evidence from Parrot Sabini and Silver (1988) who show very clearly how this happens in a personal relationship. They found their subjects did not feel comfortable saying something unpleasant to someone else. Baron and Byrne (2000 p. 364) argue, in many situations if individuals feel uncomfortable or ill at ease they tend to try to conform or fit in with others. In a sensitive situation individuals will try to react in a manner which does not upset other people. Further support for these findings comes from the work of Ybarra (1999) who indicated people like to feel good about themselves. They are particularly reluctant to be negative about other people. These findings can be applied to the manner in which evaluation forms are completed and would suggest those who fill in the forms may be reluctant to be negative about people involved in the courses.

A study by Wall (1973), found we tend to respond positively about other people, and feelings such as being fulfilled, but that participants who were found to have high levels of social desirability were more likely to respond negatively about inanimate things such as the working environment. This work was conducted in industry but if applied to course evaluations would suggest an individual completing an evaluation would react differently when feeling positive than when feeling negative and that a desire to react in a socially desirable manner may be responsible for this.

Evaluations and the training environment
Not all ‘How to do it books’ on evaluating, such as that written by Rae (2002) draw attention to the environment, in which courses are held. There is invariably a domestic side of courses, such as the provision of coffee, room comfort, visual aids and general administrative arrangements. These are all factors which are classified in this article as ‘hygiene factors’ following Herzberg (1966). Over many years it has been become clear to the researcher that course participants react to these sort of factors. This has been supported by researchers such as Andersen et al (1999 pp 359-374), who stress the need to provide a supportive and comfortable environment for training sessions.

Transfer of learning
The present study examines evaluations of courses immediately after the courses were completed. What is being judged is the conduct of the course itself. This aspect of course evaluations has been neglected in recent research as attention has been directed more towards evaluating the transfer of learning which usually involves assessing the extent the participants have adopted aspects of the course into their working practice. Evaluating transfer of learning has tended also to concentrate on aspects of training such as, goal orientation (Colquitt et al 2000), motivation on training outcomes (Tracey et al (2001), the presence of desired rewards (Goldstein and Ford 2002) and the tradition within organisations of a learning culture (Chiaburu and Tekleab 2005). The present study which is concerned with the participants’ view of the course itself and in the selection of the courses to be evaluated has taken many of these variables into account as will be explained later. The evaluation of any course immediately it has been completed is important for any trainer. The favourable or unfavourable reaction to a course by the student can have an influence on any decision either to reject or to adopt and transfer to the workplace what has been learned. This initial evaluation of the reaction to a course is the subject of this study.
Aim of the study
This study aims to examine whether a desire to respond in a socially desirable manner influences responses on open ended evaluations of training courses. To do this open ended evaluations on two large training programmes were examined. Warr’s reference of the perception of personal factors in the work setting to the training situation provides the basis for one of the categories of response. References to the presenter as a person, the style of presentation used, the interaction between presenter and course participants and the interaction between course participants themselves, course content, groupwork and individual work are, for convenience, referred to in this paper as ‘human related factors’. Following Herzberg (1966) and Andersen et al (1999 pp 359-374) references to the training situation including the comfort and ambience of the training room, administrative factors such as joining instructions, the catering facilities, the visual aids, the library and other informational resources are referred to in this paper as ‘hygiene factors’.

METHOD
Two hypotheses were tested:
Hypothesis 1: Elements of the courses concerning ‘human related factors’, will be more likely to receive favourable than unfavourable comments.

Hypothesis 2: Elements of the courses concerning ‘hygiene factors’ will be more likely to receive unfavourable comments than favourable ones.

The participants
Two training programmes were used in this study. One concerned ways of dealing with child abuse in school. The total recruitment on this programme was 377 senior teachers, deputy head teachers and head teachers. This course was repeated 17 times with class sizes of about 25. It consisted of a two day session and a further one day session held after an interval of a few weeks, both of which were evaluated.

The second training programme was a three half day teaching skills course. The total recruitment on this programme consisted of 231 postgraduates from many different academic disciplines, and nearly a quarter of them had English as a second language. This course was repeated 25 times with class sizes of between 8 to 10 and each half day was held on consecutive weeks. This was evaluated at the end of the third session.

The choice of courses explained
Evaluation studies involving college students and those investigating the transfer of learning can generally be criticised for failing to control a number of variables. In the present study the two courses were selected as they enabled many of these to be taken into account. First all participants were taking part as a requirement of their jobs and so the variable of participants choosing to take part was eliminated. Pohlmann (1975) found student’s evaluations of ‘electives’ tend to be far more favourable than those of ‘required’ courses. In the present study courses were ‘required’. The participants were not seeking promotion or advancement from their performance on the course which is an issue noted by Tracey at al (2002), but were merely learning information and skills which would help them with their existing duties. Both courses concerned skills which were outside their past experience and so the variable of prior knowledge or attitudinal prejudices was kept to a minimum. Both courses involved both interpersonal skills and skill learning. The very different characteristics of the participants, in terms of age, status and academic background on the two courses was intended to demonstrate that any results could be applied to more than one population group. A variable noted by Chiaburu and Tekleab (2005) was taken into account as the academic level of the two sample populations was specifically chosen. This was to ensure that they had experience of participating in training courses, were familiar with a culture of training, and hence would have experience of standards of presentation against which to compare the courses. Greenland and Gillmore (1997) are typical in their findings that students who obtain a high grade tend to regard a course more favourably than students, who obtain a low grade. No grades were awarded for either programme included in the present study and no reward was given for Goldstein and Ford (2002) have identified this as an intervening variable. Variations in class size have been shown to have an impact on evaluations of courses, although, according to Scott (1977) the direction of impact is not clear. In the present study group sizes were kept constant. According to Bassin (1974) a weaknesses of studies which include college students is that rarely is the same course run more than once without the content changing. It may be that the lecturer changes the information given, or someone else teaches the course. This results in rarely more than one or two intakes of student evaluating the ‘same’ course. Both the programmes in this study were designed at the outset and as far as was practicable content was kept the same each time they were repeated, over a two year period. Many intakes of students were thus able to evaluate them.
Course content and teaching methods:
Each course included information input and certain practical skills which needed to be acquired. Teaching methods employed in both courses were based on a ‘workshop’ approach. This involved presenter input, group work and individual exercises.

Open ended evaluation employed on both courses:
For the child abuse course the trigger words were ‘hindered’, ‘helped’ and ‘other comments’. Every participant was given a sheet divided into three sections each headed by trigger words. The actual form was a single side of a A4 sheet. The trigger words used for the teaching skills course were: ‘The best thing about the course’, or ‘the worst thing about the course’. Ninety five per cent of students returned forms of which less than one per cent. were returned blank.

In addition to the open ended response forms the participants on the postgraduate teaching skills course were given a structured ‘Likert’ type scale evaluation. This was included in the present study as a control for the open ended responses. Any response on the open ended form may well be influenced by social desirability. One way of demonstrating this has occurred is if the ‘Likert’ response to a similar negative statement is more or less extreme. Kobrynowicz and Biernat (1997), in a study comparing open ended and ‘Likert’ style response forms, have shown open ended response forms allow for a greater degree of expression than structured ‘Likert’ style response forms.

Categorisation of responses:
This involved allocating each individual statement, on every open ended evaluation form to a category. This was done thematically, initially using a hypothetico-deductive approach (Hayes 2000 p 179). and then an inductive approach was adopted. The thematic analysis was based partly on the initial literature review which highlighted ‘human related factors’ and ‘hygiene factors’. Parrot Sabini and Silver (1988), for example, stressed the importance of positive reactions to persons and the use of inanimate areas to express negative views. Morgan, Carder and Neal (1997) stressed the importance of groups and how we turn to them for support. Further factors were derived from Herzberg (1966) who referred to feelings of achievement and satisfaction which people express when they are feeling positive about their work. References to the presenter as a person, the style of presentation used, the interaction between presenter and course participants and the interaction between course participants themselves, course content, groupwork and individual work are referred to subsequently in this article as ‘human related factors’. Herzberg (1966) and Wall (1973) highlighted the use of ‘hygiene factors’ when wanting to express displeasure. In the present study references to the training situation including the comfort and ambience of the training room, administrative factors such as joining instructions, the catering facilities, the visual aids, the library and other informational resources were categorised as ‘hygiene factors’.

The statements on the evaluation forms, from both courses, were categorised by the researcher. Twenty per cent were selected at random by an assistant, who was instructed in the categorisation scheme and totally independently scored 120 forms. 420 individual statements on these forms were placed in categories and the assistant allocated 382 in the same categories as the researcher. This was a 91% matching rate.

Scoring the open ended evaluations:
The evaluations of the participants were scored according to the order of the comments made. The ‘helped’ or ‘best’ comments were recorded separately from the ‘hindered’ or ‘worst’ comment. For each the first comment made was awarded a score of four the second comment was awarded three, the third comment was awarded two, and the fourth and subsequent comments were awarded one. When no comment was made that category was awarded zero. The positive and the negative responses were scored separately but both were scored in the same way, so the first positive comment made would be scored 4 and the first negative comment would also be scored 4. This took into account the order effect (eg Sherman and Klein, 1994; Swann and Gill, 1997; Wyer et al, 1994).

Categorisation of Likert scale responses with teaching skills course:
For purposes of this study the questions on the Likert scale were also allocated to either ‘human related’ or ‘hygiene factors’ by five raters acting independently. Four ‘Likert’ style statements referred to ‘human related factors’ and four to ‘hygiene factors’, as can be seen in table I. Each was scored out of 5 but averaged for the analysis. This gave a minimum score of 1 and maximum of 5.
Table I
Showing the questions included in the Likert scale evaluation. Categorisation based on the scheme used for the open ended evaluation.

<table>
<thead>
<tr>
<th>‘Human related statements’</th>
<th>Very poor</th>
<th>Poor</th>
<th>Average</th>
<th>Good</th>
<th>Very good</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality of presentations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quality of group management</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Integration of parts</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Appropriateness of level</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| ‘Hygiene statements’       |           |      |         |      |           |
| Consistency with publicity |           |      |         |      |           |
| Quality of audio-visual    |           |      |         |      |           |
| Quality of handout materials |        |      |         |      |           |
| Followed good equal opportunities practice | | | | | |
| Efficiency of course administration | | | | | |

The analysis of the data:
In order to test the hypothesis, the positive and negative scores were compared for the ‘human related factors’. This was to see whether positive and negative scores were significantly different. The positive and negative scores for the ‘hygiene factors’ were also compared.

RESULTS
Hypothesis 1: Elements of the courses concerning ‘human related factors’, will be more likely to receive favourable than unfavourable comments.
The t-test results shown in table II support the hypothesis. The positive evaluations of the ‘human related factors’ are higher than the negative ones for all three evaluations. These are the child abuse groups after days one and two, again after day three and also the postgraduate teaching skills group.

Table II
Showing t-test results for related samples comparing positive and negative scores for the ‘human related’ categories

<table>
<thead>
<tr>
<th>Group</th>
<th>Means. Standard deviations in brackets</th>
<th>t. df. Probability (2 tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Positive</td>
<td>Negative</td>
</tr>
<tr>
<td>Child abuse days one and two</td>
<td>6.09(2.00)</td>
<td>1.93(2.13)</td>
</tr>
<tr>
<td>Child abuse day three</td>
<td>5.43(2.00)</td>
<td>1.96(2.00)</td>
</tr>
<tr>
<td>Postgraduate teaching skills</td>
<td>4.41(2.06)</td>
<td>2.15(2.01)</td>
</tr>
</tbody>
</table>

Hypothesis 2: Elements of the courses concerning ‘hygiene factors’ will be more likely to receive unfavourable comments than favourable ones.
The analysis supports the second hypothesis. The negative ‘hygiene’ scores are higher than the positive ‘hygiene’ scores for all three of the evaluations as can be seen in table III.

Table III
Showing t-test results for related samples comparing positive and negative scores for the ‘hygiene factors’

<table>
<thead>
<tr>
<th>Group</th>
<th>Means. Standard deviations in brackets</th>
<th>t. df. Probability (2 tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Positive</td>
<td>Negative</td>
</tr>
<tr>
<td>Child abuse days one and two</td>
<td>0.35 (1.38)</td>
<td>0.76(1.55)</td>
</tr>
<tr>
<td>Child abuse day three</td>
<td>0.07 (0.48)</td>
<td>0.67(1.46)</td>
</tr>
<tr>
<td>Postgraduate teaching skills</td>
<td>0.08 (0.51)</td>
<td>0.28(1.01)</td>
</tr>
</tbody>
</table>
With the findings for the child abuse courses it is possible, although very unlikely, that all the presenters were good and the venues and audio visual aids, administration etc were poor, rather than that the participants were responding in a socially desirable manner. With the postgraduate teaching skills course this possibility was taken into account. In addition to the open ended response forms the participants were given a structured evaluation using a Likert style scale, which specifically mentioned many of the ‘human related’ and ‘hygiene factors’ included by participants in the open ended response forms. According to other researchers such as Kobrynowicz and Biernat (1997) Likert style response forms allow less freedom of expression than open ended ones. As can be seen in table IV there was no significant difference between reactions to ‘human related’ and ‘hygiene factors’.

Table IV
Showing t-test results for related samples for the Likert style questionnaire completed by the postgraduate sample on the teaching skills course, comparing ‘human related’ and ‘hygiene’ overall scores

<table>
<thead>
<tr>
<th>Group</th>
<th>Means scores with Standard deviations in brackets</th>
<th>t.</th>
<th>df</th>
<th>Probability (2 tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Postgraduate Teaching skills</td>
<td>Human rel. Hygiene</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.28 (.53) 4.23 (.50)</td>
<td>1.66</td>
<td>230</td>
<td>.099</td>
</tr>
</tbody>
</table>

Both ‘human related’ and ‘hygiene’ scores were above 3 which was the mid point on the scale. The correlation between the two was .53, which is in the medium range (.30 to .70), which Sheehan and DuPrey (1999) consider to be of moderate significance. Below .3 they regard a correlation as of no interest and above .7 to be extremely interesting. This indicates on a scaled response form that the participants at the time of completing the open ended forms, regarded the ‘human related’ and the ‘hygiene factors’ similarly. This is not the impression created on the open ended forms. This would suggest something is influencing responses on the open ended evaluations. This strongly suggests the ‘hygiene factors’ were not really regarded unfavourably but that in the open ended statements participants, it is argued here, seem to be responding in a socially desirable manner.

It is clear from the open ended data, many more comments are made about ‘human related’ than ‘hygiene factors’. This indicates the far greater salience of the former to the course participants. This is not demonstrated in the Likert scale reported in table IV. This does illustrate one of the limiting characteristics of a Likert scale, which does not provide an indication of the salience of various items to those completing it.

A further check on the ‘social desirability’ interpretation:
To provide further support for the interpretation that the results of the open ended responses are influenced by social desirability, twenty undergraduates aged 18-47 (mean 33 years), non of whom were involved in either training programme, were given a questionnaire. With reference to the last course they had attended they were asked, which from a list of factors presented to them individually on separate cards, they would feel most comfortable criticising negatively.

The actual instructions for completion of the task were as follows:
‘I would like you to think of the last course you have completed. Rank the items on the cards in terms of how comfortable, in a socially acceptable sense, you would have found it to have made unfavourable or negative comments about the following features of the course. Rank order the cards, giving the most uncomfortable a score of one and placing it at the top of the list.’

The cards were laid out in random formation on a table. Each card contained details of one possible evaluation response. This follows a well established experimental procedure. Holmes and Rahe (1967), for example, used this method to design a stress scale. Many researchers have used this procedure since that time to establish various scales.
Table V

Showing the average score allocated to each statement from the 20 participants. The lower the score the more uncomfortable the participants would have felt. ‘Human related factors’ are marked HR. ‘Hygiene factors’ are marked H.

<table>
<thead>
<tr>
<th>Statements</th>
<th>Mean score, standard deviation in brackets</th>
</tr>
</thead>
<tbody>
<tr>
<td>The presenter’s style of presentation HR</td>
<td>2.15 (2.37)</td>
</tr>
<tr>
<td>The atmosphere created by the presenter HR</td>
<td>2.15 (0.75)</td>
</tr>
<tr>
<td>The knowledge of the topic of the presenter HR</td>
<td>2.65 (0.93)</td>
</tr>
<tr>
<td>The presenter’s group management skills HR</td>
<td>4.10 (1.16)</td>
</tr>
<tr>
<td>The other students on the course HR</td>
<td>5.15 (0.93)</td>
</tr>
<tr>
<td>The content of the course HR</td>
<td>6.45 (1.67)</td>
</tr>
<tr>
<td>Taking part in course activities HR</td>
<td>7.55 (1.50)</td>
</tr>
<tr>
<td>The audio visual aids H</td>
<td>8.90 (0.55)</td>
</tr>
<tr>
<td>The classroom facilities H</td>
<td>9.75 (0.91)</td>
</tr>
<tr>
<td>The refreshments provided H</td>
<td>11.00 (0.86)</td>
</tr>
<tr>
<td>The comfort of the room H</td>
<td>12.35 (0.67)</td>
</tr>
<tr>
<td>The administration H</td>
<td>13.05 (0.60)</td>
</tr>
<tr>
<td>The car parking H</td>
<td>12.60 (3.10)</td>
</tr>
</tbody>
</table>

From the results shown in table V it would seem students feel less comfortable criticising ‘human related factors’ than they do ‘hygiene factors’. This provides further support for the view social desirability has a part to play in how open ended evaluation forms are completed.

DISCUSSION

These findings provide a lot more information about open ended evaluations than is provided in the research methods texts (eg Bryman 2001, Erickson and Kaplan 2000, Jackson and Trochim 2002 and Sommer and Sommer 1997), which aim to outline the advantages and disadvantages of various evaluation methods. The major finding reported concerns the use of open ended evaluations. One argument for using open ended evaluations put forward by researchers such as Jackson and Trochim (2002), is it enables course participants to respond freely and express things about the course which concern them. They suggest they may not be able to do this within the confines of a structured questionnaire. This argument does not take into account factors which have been highlighted in this study. The results of the open ended evaluation reported here immediately the courses were finished, showed participants did not appear to be totally free in their comments. They performed as though they wanted to see themselves in a socially acceptable light. They praised presenters and other ‘human related factors’ and criticised ‘hygiene factors’ which, they felt were ‘not personal’. This finding was based on a study which included two sample populations, one of child abuse co-ordinators and another of post graduate research students attending a teaching skills course. This certainly is a very different view to the one presented by Leslie Rae (2002) when he suggests the learners are in control of the evaluation comments they make when the open ended format is used. It is argued here what they said about the course was influenced by a desire to respond in a socially desirable manner.

When students mentioned the presenters or course contents they tended to respond positively. When issues such as the venue or catering are referred to (Herzberg’s 1966 hygiene factors) students tended to react negatively. The interpretation offered in this thesis was inspired by Wall’s (1973) explanation in terms of sociable desirability. In order to express unfavourable feelings, they selected an issue which was not personal, or socially unacceptable. This takes our understanding of the influence of social desirability further than the simplistic notion offered by Diamond (1972) who merely noted we respond in a socially desirable manner and did not make any distinction between positive and negative reactions.

It was originally mentioned researchers such as Memon and Bull (2000) have shown when interpreting responses in an interview, which is obviously not anonymous, account needs to be taken of attempts to create a favourable impression. The idea that open ended questions provide an opportunity for course participants to express themselves freely which is promoted by researchers such as Jackson and Trochim (2002) is brought into question by the results of this study. It is assumed participants say what they want, particularly when, according to Diamond (1972), they cannot be identified. This study has shown they can
be influenced by psychological factors such as the desire to see themselves in a socially acceptable manner, even though their responses were anonymous. It would appear from evidence presented here respondents are, possibly without realising it, trying to create an impression by reacting in a socially desirable manner. Whoever interprets evaluation responses needs to take this into account.

It was found that although open ended evaluations may have some limitations for favourable reactions, they may have even more limitations when unfavourable reactions are considered. When being negative, students said they felt more comfortable commenting on the ‘hygiene factors’. It would not be expected that the presenters (one of the ‘human related factors’) really were all liked, and the venues (one of the many ‘hygiene factors’) really were all very poor. It is clear from responses on the teaching skills Likert style evaluation scale this is not the case. On the Likert scales the ‘human related’ scores are not significantly different to the ‘hygiene’ scores. The characteristics of the two courses provide further support for this interpretation. The interpretation that the presenters (‘human related factors’) were good and the venues (‘hygiene factors’) poor is not acceptable. There were 22 presenters involved with the two courses. It would be unusual if all of these were seen by the majority of course members as being excellent. Second there were a total of 19 different venues used for the two courses. Again it would be unlikely if all were regarded negatively. Indeed, in an attempt to compensate for the compulsory nature of the courses, efforts were made to ensure the venues, visual aids and catering arrangements were better than was customary for the courses typically attended by the participants.

These results indicate care should be taken when interpreting open ended evaluations. This means organisations need to compare the evaluations of one course, with that of others rather than just assume the presenters are good because they receive favourable comments on open ended evaluations. It appears the participants are willing to praise individuals in various ways, but they are not too willing to criticise them. They are reluctant to cause distress and will find another factor against which to direct their complaint. It is suggested, in this study, they are likely to direct it against ‘hygiene factors’. They find a socially accepted direction to place their criticisms. Any interpretations made from open ended evaluations need to take this tendency into account or very false conclusions may be drawn about the merits of presenters and the inadequacies of the venue. These results suggest that the providers of the venues (which are one aspect of the ‘hygiene factor’) may be subject to far more criticism, than is justified. Vast sums have been spent by universities, colleges and hotels providing state of the art conference venues, yet from the evidence of this study, it is doubtful if responses on open ended evaluation forms will reflect this favourably.

These findings do suggest those who use open ended evaluations need to be particularly careful when they interpret them. They need to be aware how the responses are influenced and that the respondents are not really expressing their free and totally honest feelings. Open ended responses are a useful means of evaluating courses, as they can provide instant feedback, which is particularly useful when reviewing a new course, lecturer or venue but responses need to be judged in the light of other evidence which places these variables in context. Comparison with the comments made about other presenters should be an absolute minimum requirement when evaluations of a single tutor are considered. It is also essential to make comparisons of comments made about course venue and hygiene factors with those made by participants attending other courses in the same venue. Most importantly those interpreting open ended evaluation comments need guidance and training in order to carry out any analysis.

Acknowledgement:
This empirical study is integrated in the author’s own doctoral thesis.

REFERENCES


**Biographical note:**

**Jenny Darby** has a varied career in Education. She taught in an inner city Comprehensive school for 25 years completing her service as Head of Science. For two years during a secondment she was responsible for a county wide training programme for head teachers and senior teaching staff. She has recently been running training workshops in teaching skills for postgraduates at Loughborough University, where she completed her doctorate. Her main research interest concerns evaluation of training programmes. E mail J.A.Darby@lboro.ac.uk