SAFETY SCIENCE Monitor

Issue 2 2009

Article 1 VOL 13

DEVELOPMENT OF AN EVALUATION INSTRUMENT TO PREDICT EFFECTIVENESS FROM TRAINING IN OCCUPATIONAL HEALTH AND SAFETY

K. MASUHR,
D. WINDEMUTH,
E. TASKAN-KARAMUERSEL

Department Evaluation of Preventive Measures, BGAG Institute Work and Health of the German Social Accident Insurance (DGUV)

ABSTRACT

Continuing vocational education in the field of health and safety aims to influence the attitude towards, the knowledge about as well as the health and safety behaviour of trainees. To reach this high educational quality of training an enormous effort is expended. To prove if this effort is expended in the most effective way, more and more educational institutions are engaged in evaluating this ratio of costs and benefit.

A lot of research was carried out by the Institute Work and Health (BGAG) the last years to prove the quality of training and to prove the transfer of training that takes place.

Quality of training implies more than the participant's satisfaction with the training. There should also have taken place learning, knowledge and skill enhancement. The learner should be enabled to perform the transfer of training into workplace.

To prove this quality of training an evaluation instrument has been developed.

A search regarding questionnaires for the evaluation of seminars and training was carried out. On the basis of 37 questionnaires from private and public educational institutions an item pool was compiled. This pool of 1024 items was structured by the content character of the item such as belonging to a certain scale, quality criteria and heritage from a certain questionnaire.

A criteria based reduction process of several rounds was carried out to eliminate items that were not fitting by content or were identical.

This reduction process leads to a questionnaire of 9 scales with all over 59 items. A following statistical testing including factor and item analysis enables further reduction of items for a revised questionnaire.

The questionnaire was developed to measure the quality of training and to predict transfer from training and herewith the long-term effect of training. Studies to prove this predictable validity will be realised in the near future.

1. QUALITY MANAGEMENT AND EVALUATION IN PREVENTATIVE SAFETY TRAINING

The institutions responsible for statutory accident insurance and prevention in Germany offer a range of prevention services to ensure compliance with the legal obligation to prevent work-related accidents, injuries and health risks. These predominantly serve to initiate and support prevention in the workplace (Wetzstein und Lauterbach, 2007). Occupational health and safety training is one of the key prevention services. This includes providing occupational health and safety training for employers, health and safety professionals, company safety specialists etc. Training seminars equip people with an understanding of the technical issues, legal framework and occupational health and safety procedures relating to specific industries. Participants acquire new knowledge and skills which they can then apply in their own workplace.

The quality of training seminars in the field of occupational health and safety must be guaranteed to ensure their success. In this context, quality can generally be understood to mean "the degree in which a set of inherent characteristics fulfils requirements" (DIN EN ISO 9000: 2000), but also "what can be improved through a constant process of monitoring and adjustment" (Jaster, 1997).
Donabedian (1966) distinguishes three types of quality:

- structure quality
- process quality
- outcome quality

It is assumed that high outcome quality can be achieved by ensuring good structure and process quality.

These three quality categories can also be applied to education and training. For example, structure quality can refer to the material, spatial and personnel dimensions of the course design. Process quality includes the actions, behaviour and procedures of the people involved in the training process. Outcome quality relates to the impact of the training and in particular the impact on the course participants, the course providers and also on the participant's place of work.

Quality management focuses on the planning, control, assurance and improvement of quality, in this case quality of training (Stockmann, 2006). Quality must be measured before it can be planned, controlled, assured and improved. This can be achieved by gathering information about the existing quality of the elements that make up the training system.

Evaluation is a method of obtaining and assessing information. Evaluation research refers to "the systematic use of socio-scientific research methods to evaluate the design, configuration, implementation and benefit of social intervention programmes" (Rossi, Freeman & Hofmann, 1988). Evaluation is therefore "the systematic acquisition of criteria-based assessment and evaluation of data relating to documents, actions and individuals for the purpose of making further decisions (Maritzen, 1996)".

When measuring the quality of training seminars, it is therefore important to establish in advance what is important in terms of quality and which aspects require examining. Quality indicators can then be compiled on the basis of which data acquisition and assessment can take place.

2. INDICATORS TO MEASURE THE QUALITY OF THE TRAINING PART OF THE PREVENTION SERVICE

As part of the "quality in prevention" research project, the sub-project "indicators of quality measurement" established a set of quality indicators for the prevention services provided by the statutory accident insurance and prevention institutions in Germany. The aim was to create appropriate indicators to assess the structure, process and outcome quality, which would then provide proof of effectiveness and efficiency. With the aid of extensive literature research, objectives were determined and indicators were defined as characteristics for all three quality categories (cf. Donabedian, 1966) across all the prevention services.

Targets were also defined for the "training" prevention service and appropriate indicators were determined to measure target achievement (Table 1).

Table 1: Targets and indicators for training seminar quality (Wetzstein and Lauterbach, 2007)

Level	Indicator
Structure quality	
	Personnel resources
	Material resources
	Availability/access
	Topicality
Process quality	
	Communication (style)
	Target-group orientation
	Transparency/standardisation
	Documentation/evaluation
Outcome quality	
	Customer satisfaction/acceptance
	Attitude change
	Mastery of new skills
	Complaint management

In this instance, *structure quality* covers personnel resources in terms of the instructors' competence, their professional qualifications, and experience with the target group and teaching qualifications, and also the competence of the seminar providers to publicise and run the seminar. Material resources are another important aspect of structure quality. The suitability of the seminar rooms and practice areas for promoting learning, a pleasant environment and the practical application of occupational health and safety provide information about structure quality. The availability and accessibility of a seminar are determined by a defined number of participants and adequate entry requirements. Another important aspect for structure quality is whether customers are satisfied with the interval between applying for the seminar and attending it.

Process quality includes clear communication between instructors and participants. For example, the conversational style and communication techniques and media used all provide information about the quality of communication. By making use of participants' previous experience and focusing on real circumstances in the workplace, the instructor can relate well to his target group (target-group orientation). The instructor can increase transparency and standardisation by formulating teaching/learning objectives. Documentation and evaluation can be used to measure, assure and control process quality, participant satisfaction and learning and transfer success. A dedicated evaluation concept can steer these processes.

Outcome quality covers customer satisfaction, in other words how satisfied a participant feels after the seminar, and attitude change. A particular goal of occupational health and safety is to induce attitudes to occupational health and safety as a condition for behavioural change. In terms of outcome quality, it must be possible to make predictions about the extent to which commitment to occupational health and safety and motivation to protect others changes. In addition, actual behaviour and mastery of new skills must also be measured after a seminar. This can be defined in terms of three levels. The first level is successfully reached when the participant tries to apply what he has learned after the seminar and has extended his repertoire of strategies as a result of attending the seminar. This second level is reached when the seminar participant behaves in a way which promotes occupational health and safety and also tries to influence the behaviour of others (e.g. colleagues). Success at the third level is attained when, in addition to the participant's behaviour, health and safety conditions in the workplace change as a result of performing and documenting health and safety-related activities.

These indicators are essential for determining and assessing the quality of health and safety training seminars. In the context of seminar evaluation and quality assurance, they have been used to devise a seminar evaluation questionnaire to measure partial aspects efficiently.

3. DEVELOPING A TOOL TO MEASURE ASPECTS OF STRUCTURE, PROCESS AND OUTCOME QUALITY

Problems and objective

The seminar evaluation questionnaire was developed within the BGAG Institute Work and Health. The BGAG is one of three important education and training institutes of the German Social Accident Insurance (DGUV). One of the institute's main functions is to provide training, in addition to research and advisory projects. The seminars build on the BGAG's training principles and feature activity, experience and skill-based criteria (Bollmann, Gallenberger, Jagenlauf and Kici 2005).Up until now the seminars have been evaluated on a regular basis using a seminar evaluation questionnaire. The content of the questionnaire has been redesigned for three reasons:

a) Results from analysing the "old" seminar evaluation questionnaire

According to the changes in the world of work new demands for the qualification of employees in occupational safety and health have been coming up. This led to adaptation of the qualification process in the BGAG in the last years. Therefore the accordance with these new requirements as well as the quality of the questionnaire were examined. This revealed that instead of recording a range of different aspects, the questionnaire recorded only three different assessment items. Using statistical analysis, it was possible to show that the first 10 questions measured only one aspect of the seminar; satisfaction with the seminar content.

b) Results from the "quality in prevention" research project

As part of the "quality in prevention" research project, the sub-project "indicators of quality measurement" identified a set of indicators for measuring the quality of the training seminars. These indicators provided guidance for selecting the content of the redesigned seminar evaluation questionnaire. Furthermore, the results from the "training" sub-project (Gallenberger, 2007) showed that transfer in particular can be largely predicted by the transfer motivation of the participant at the end of the seminar. The project also revealed that after a seminar 66.1% of participants reported that they would consult the seminar handouts when applying what they had learnt during the seminar. This indicates that particular emphasis should be placed on well-designed, intelligible seminar handouts.

c) Measuring quality indicators at the level of structure, process and outcome quality

The seminar evaluation questionnaire was intended to measure aspects at all three quality levels (structure, process and outcome, see Table 1) and provide the BGAG and instructors with information which can be used for quality control.The aim is to devise a seminar evaluation questionnaire which can be used to assess all BGAG training seminars for safety specialists, inspectors, instructors, technical specialists, managers, occupational health physicians and medical auxiliaries.

The seminar evaluation questionnaire should perform the following functions:

- acquisition of information for instructors and training provider (BGAG)
- quality control
- guidance to improve seminars
- predictions about knowledge transfer.

The seminar evaluation questionnaire provides feedback about seminar quality at the end of a seminar. Aspects of structure and process quality in particular, and to some extent outcome quality as well, should be measured. The seminar questionnaire is to be used routinely after each seminar and is a particularly important tool, since it provides the only opportunity to regularly control and manage seminar quality. Consequently the quality and suitability of the questionnaire for acquiring and evaluating information must be regularly monitored and adjusted as required.

QUESTIONS

The questionnaire should satisfy scientific quality criteria and criteria for devising questionnaires. In addition, the questionnaire is required to provide information about relevant aspects of seminar quality. The studies being carried out should focus on the following questions:

1. What is the value of the items in the questionnaire being devised?
2. What factor structure emerges on the basis of the items for inclusion in the questionnaire?
3. Can item analysis provide guidance for reducing the number of items in the questionnaire?
4. Is it possible to devise a questionnaire which can be used to measure several dimensions of seminar quality?
5. Can the results be replicated?

SAMPLE

Seminars for safety professionals, inspectors, instructors, technical specialists, managers, occupational health physicians and medical auxiliaries at the BGAG were evaluated from January to April 2008. In the second study, seven further seminars provided by another DGUV training provider were evaluated.

In the first study, 148 people evaluated 13 BGAG seminars using the first version of the seminar evaluation questionnaire. In the second study, 14 seminars were evaluated by 223 people using the simplified version (N) of the second questionnaire. 146 people evaluated 12 seminars using the version of the questionnaire with more extreme phrasing (S).

METHODS

Document analysis was carried out to select items for the first version of the seminar evaluation questionnaire. Internally and externally devised questionnaires for evaluating seminars and other courses were examined. Criteria-led item reduction resulted in the first version of the seminar evaluation questionnaire.

Item analysis was carried out with the aim of evaluating the quality of items included in the questionnaire and obtaining guidance on item reduction. Item analysis involved determining the index of difficulty, factor structure, discriminative power and Cronbach's alpha for the individual scales and the whole questionnaire. Item difficulty is measured by means of an index. This difficulty index or popularity index enables predictions to be made about the levels of agreement per item (Bortz und Döring, 2002). Factor analysis indicates which dimensions can be taken as a basis for the items and which items can be combined to form scales based on similarity of content. The item discrimination index provides information about how representative an individual item is of the overall result of a scale.

The results of item analysis provide information about item reduction and therefore re-designing the questionnaire.

RESULTS

Literature research

a) Study of evaluation questionnaires for seminars and other courses

To develop the seminar evaluation questionnaire, 37 standardised questionnaires designed to evaluate continuing vocational education courses were analysed. Four questionnaires developed by the BGAG were also included in the study. All the items from the questionnaires were combined to form an item pool of 1024 items. The relevant scales, number of answer levels and information about the quality criteria for the items and scales were given in addition to the formulated items. At the same time, the item pool was arranged according to content.

b) *Criteria-led item reduction*

The first criterion for item reduction was item quality. Poor quality items were removed from the item pool. The second criterion concerned university seminars; since questionnaires evaluating university courses and seminars were also included in the item pool, items relating to specific studies and to university education were ignored. The third criterion concerned the relevance of the items to BGAG seminars. Items were selected on the basis of compliance with the BGAG's training principles (cf. Bollmann et al.), and they were also required to be indicators of training quality; the fourth criterion. A questionnaire was compiled using the remaining 59 items and submitted to training specialists at the BGAG for evaluation.

c) Compiling a preliminary version of the questionnaire

Following item reduction, the 59 items were combined in a six-point Likert scale to create the first version of the seminar evaluation questionnaire. The items can be arranged by content in the following sections:

-organisation
-seminar content
-practical relevance
-interaction
-learning climate
-seminar handouts
-seminar satisfaction
-learning success
-transfer motivation.

Questions in the organisation section cover structure quality in terms of the course providers' ability to organise and run the seminar. Insights about the process quality of the seminar can be gained by evaluating seminar content, practical relevance, interaction during the seminar, learning climate and seminar handouts. The satisfaction, learning success and transfer motivation sections, on the other hand, cover aspects of outcome quality.

STUDY 1: ITEM ANALYSIS AND REDUCTION

Item analysis of the questionnaire was performed on the basis of 148 datasets.

Item difficulty index

Calculating the difficulty indices, in other words the levels of agreement between respondents, showed that on the whole individual items were evaluated very positively. Almost all respondents assess items with a difficult index below 0.2 as not applicable and above 0.8 as highly applicable. A strong acceptance or rejection of an item provides little information for evaluating seminars. The difficulty index therefore provides guidance on which items should be removed from the questionnaire.The calculated difficulty indices for the seminar evaluation questionnaire ranged from 0.59 to 0.94, with 27 items exceeding the critical value of 0.8.

Factor analysis

Factor analysis was performed with orthogonal rotation (Varimax). Parallel analysis was used as a criterion for factor extraction. Hereby as many factors are extracted, as their Eigenvalues are larger than coincidentally determined Eigenvalues. A factor analysis corresponding to the principal component method yielded a five-factor solution, which accounted for 61.23% of the total variance. At this stage all items with a too low factor loading (< 0.5) or no clear factor loading had already been eliminated. Subsequent interpretation of the factor solution in terms of content revealed that the questionnaire is based on the following five dimensions, according to which a BGAG training seminar can be evaluated:

1. Seminar content
2. Interaction within the seminar
3. Quality of seminar handouts
4. Transfer motivation of participants
5. Seminar organisation.

Reliability and discriminative power

Factor analysis provides information about which items can be combined to form scales. To further determine item quality, the item discrimination index of individual items was then calculated on the basis of these scales.
This resulted in a very broad spectrum with item selectivity ranging from scales of .44 to .89. The values indicate the extent to which an item represents the characteristic measured by the scale and also assist with item reduction.

Cronbach's alpha was also calculated for the scales and the whole questionnaire. Values ranged from .76 - .96 for the individual scales and the total value (Table 2). It shows that the individual scales measure each characteristic very well, although the total value as a measurement of the "seminar success" characteristic is somewhat higher.

Table 2: Reliability of Version 1 for the scales and the whole questionnaire

Scale	Cronbach's alpha for scale
Content	.960
Interaction	.868
Seminar handouts	.944
Transfer motivation	.918
Organisation	.758
Cronbach's alpha total	.967

Item reduction

The results of item analysis were collated in one decision matrix per item and compared. The results of the difficulty calculations, factor analysis and discriminative power calculations were integrated and a criteria-based

decision about each item was made. An inconclusive factor loading or a factor loading below .5 was used as a criterion for elimination. This procedure also resulted in the removal from the questionnaire of items with extreme item difficulty indices (<0.2 and >0.8). Consequently, 29 items were removed from the questionnaire on the basis of their statistical characteristic values. Despite good statistical characteristic values, it was clear that the seminars had been evaluated with above-average positivity (see difficulty indices). However, the questionnaire should also demonstrate potential to change and a further version of the questionnaire will be developed in the second stage.

STUDY 2: ITEM ANALYSIS AND REDUCTION TO PRODUCE A FINAL VERSION OF THE SEMINAR EVALUATION QUESTIONNAIRE

The second version was developed in an attempt to address the problem of limited utilisation of the answer scale and corresponding low variance. To this end, a parallel form of the questionnaire was developed in addition to the reduced version in the second test phase. This parallel form contains more extreme phrasing and was intended to force respondents to make a more critical judgement and therefore use the whole answer scale.

After the test phase, a total of 369 datasets were available for both questionnaires (Version 2 (N) and Version 2 (S)), each with 25 items. On the basis of this database each version was subject to item analysis to determine the difficulty indices, factor structure, item discrimination indices and Cronbach's alpha.

Item difficulty index

The difficulty indices were determined for individual items in each version of the questionnaire. Due to the positive acceptance of the items, the indices in both versions are relatively high, on average around 0.74 – 0.9 (N) and 0.65 – 0.94 (S).

Factor analysis

Factor analysis with orthogonal rotation (Varimax) was also performed in this phase. For both versions this resulted in a clear five-factor solution, which explains 76.78 % (N) and 70.27 % (S) of the total variance. Consequently, the explained variance of the second version of the questionnaire is already higher than that of the first. All items have factor loadings above .5. In terms of content, the underlying dimensions could be interpreted consistently across both versions. Factors from the first phase can also clearly be replicated:

1. Seminar content
2. Interaction within the seminar
3. Quality of seminar handouts
4. Transfer motivation of participants
5. Organisation of seminar

Reliability and discriminative power

Subsequent determination of the discriminative power for each version of the questionnaire reveals only minor differences between the two versions. Item-scale correlations (discriminative power) are between .43 - .92 for the standard version (N) and between .47 - .87 for the more explicit version (S). The Cronbach's alpha for individual scales is .77 - .96 (N) and .79 - .93 (S). For the whole questionnaire, the Cronbach's alpha is .9 for Version N, and .936 for Version S.

Table 3: Reliability of versions N and S for the scales and the whole questionnaire

Scale	Cronbach's alpha for scale (N)	Cronbach's alpha for scale (S)
Content	.889	.897
Interaction	.848	.814
Seminar handouts	.959	.931
Transfer motivation	.825	.893
Organisation	.769	.793
Cronbach's alpha Total	.900	.936

Item reduction

The results of item analysis for both versions of the questionnaire showed that both delivered very good, consistent results overall, but the results per item showed slight differences in some cases. The results were therefore used to compile a decision matrix, on the basis of which a decision was made to select each item from either the standard (N) or the more explicit (S) version.

DISCUSSION

By calculating the characteristic values of the items it was possible to show how good the individual items were and which ones were to be removed from the questionnaire. Item difficulty, calculation of factor structure, discriminative power and Cronbach's alpha provided the decision-making basis. When determining the factor structure in particular, it appeared that the factors could clearly be replicated.

Furthermore, the results also showed that the attempt to increase variance using the more extreme item formulations was not entirely successful. BGAG training seminars are evaluated very highly and changing the item formulation did nothing to change this.

It could be shown that this seminar evaluation questionnaire is scientifically based, statistically stable and can be used to measure reliably and consistently the five dimensions listed above. At the same time, the appropriate number of items (at least four items per scale) provides a meaningful feedback database for seminar instructors.

Outlook

Any further development and testing of the seminar evaluation questionnaire will focus predominantly on replicability of the item analysis results and validity determination. The questionnaire can supposedly be used to predict transfer success. In particular, it is assumed that transfer motivation and evaluation of learning content can contribute to the prediction.

Verification of criterion validity is therefore important. The corresponding external criteria are learning success and knowledge transfer. Further studies should be which relate the seminar evaluation to an objective measurement of learning success and transfer. The transfer questionnaire developed in the BGAG can be used for this purpose.

REFERENCES

Bollmann, U., Gallenberger, W., Jagenlauf, M. & Kici, G. (2005). Entwicklung von Evaluationswerkzeugen am Beispiel Berufsgenossenschaftliches Institut Arbeit und Gesundheit (BGAG), Dresden. *Grundlagen der Weiterbildung*, 16, 1, 44-51.

Bortz, J. & Döring, N. (2002*). Forschungsmethoden und Evaluation.* 3. Aufl., Berlin: Springer Verlag

DIN EN ISO 9000: 2000

Donabedian, A. (1966). *Evaluating the Quality of Medical Care.* Milbank Memorial Fund Quarterly: *Health and Society*, 44, 166–203.

Gallenberger,W. (2007).Qualität in der Prävention – Teilprojekt Qualifizierung. *http://www.dguv.de/bgag/de/forschung/forschungsprojekte/qdp/qdp_abschluss/index.jsp*

Jaster, H.-J. (1997). *Qualitätssicherung im Gesundheitswesen.* Stuttgart: Thieme.

Maritzen, N. (1996). Sich selbst und anderen Rechenschaft geben. Qualitätssicherung durch Evaluation. *Pädagogik* 48, 1, 25-29.

Rossi, P.H., Freeman, H.E. & Hofmann, G. (1988). *Programm Evaluation: Einführung in die Methoden angewandter Sozialforschung.* Stuttgart: Enke.

Stockmann, R. (2006). *Evaluation und Qualitätsentwicklung. Eine Grundlage für wirkungsorientiertes Qualitätsmanagement*. Münster: Waxmann.

Wetzstein, A., Lauterbach, D. (2007). Qualität in der Prävention – Teilprojekt Indikatoren zur Qualitätsmessung Qualitätsmessung. *http://www.dguv.de/bgag/de/forschung/forschungsprojekte/qdp/qdp_abschluss/index.jsp*