SAFETY SCIENCE Monitor

Issue 2 2013 ISSN 1443-8844

Article 4 VOL 17

MEASURING THE EFFECTIVENESS OF TRAINING IN OCCUPATIONAL HEALTH AND SAFETY

MARIA KLOTZ
Institute for Work and Health of the German Social Accident Insurance

ABSTRACT

Background: Assuring quality and measuring effectiveness are the two main functions of training evaluation. For most training providers it is difficult to find evaluation tools of good quality on the market. Very often the consequence is that there will be no evaluation at all. Some providers design their own instruments. Those are very often questionnaires of low quality, because they just randomly accumulate items for a questionnaire without a principle of choice or theory behind it. At the Institute for Work and Health (IAG) of the German Social Accident Insurance (DGUV) two evaluation questionnaires have been developed. The first questionnaire, the Seminar Evaluation Questionnaire (SEQ) looks at the quality of occupational health and safety (OH&S) training and has already been published in this journal (Masuhr et al. 2009). Herein, the quality of training is defined as the degree to which a set of inherent characteristics fulfil the requirements according to the DIN EN ISO 9000:2000. The second instrument is the Transfer Evaluation Questionnaire (TEQ). It focuses on measuring transfer which in this case is defined as putting into practice what has been learned in training. **Objectives:** The aim of this paper is to show in study 1 whether the psychometric quality of the SEQ and its theoretical framework can be confirmed with empirical data (sample of 7898 people). The purpose of study 2 is to provide psychometric measures for the TEQ and to prove whether the relationships formulated by the underlying theory can be confirmed with empirical data as well (sample of 1428 participants). Study 3 will present some results of transfer activity from participants of approx. 100 OH&S seminars given by one of the German Social Accident Insurance Institutions for the public sector. **Methods:** For the quality tests item difficulty, factor analysis, discriminative power and the reliability in terms of Cronbach's Alpha was calculated. Since the two measurement tools were developed based on well-researched evaluation models, the theoretical background was also tested by using bivariate correlations calculated with Spearman's rho. In study 3 means, standard deviations and percentages were calculated to describe the results. **Results:** The results show that the SEQ and the TEQ are of good psychometric quality and the theoretical background could be replicated with empirical data. With an ex post facto design the TEQ was put into practice and it could be shown that transfer takes place and even leads to some changes at the company level. **Conclusion:** It is possible to go far beyond the level of satisfaction by looking at training evaluation. The two measures presented show some ideas how to achieve that. Still missing are some values of reliability to show whether the questionnaires measure accurately. Also values of validity are needed to prove whether the questionnaires really measure what is claimed to be measured. By looking at the theoretical framework of the TEQ it appears that observing the last level of Kirkpatrick's (1959, 1967 & 1976) evaluation model remains a challenge. The solution of the TEQ designer is not completely satisfying and needs rework.

1. INTRODUCTION

Especially in the field of health and safety, it is not enough to just look at satisfaction of the participants at the end of training. Nevertheless, "happy sheets" are still a very common practice. It is difficult to find evaluation tools on the market which go beyond the level of satisfaction and are of good psychometric quality on top of that. Therefore the aim of this paper is to present two questionnaires for evaluating trainings which go beyond the level of satisfaction and to show, how the quality of measures can be assessed. In the two first studies presented, the

psychometric properties of the measures will be tested and in the last study some results, in terms of transfer activities from the practical field are shown. The Seminar Evaluation Questionnaire (SEQ) and the Transfer Evaluation Questionnaire (TEQ) were developed at the Institute for Work and Health (IAG). The IAG is one of three academies of the German Social Accident Insurance (DGUV). In addition to research and consulting projects, one of the institute's main functions is to provide training. Both questionnaires are based on well-researched evaluation models which will be explained in the following paragraph. The two instruments will be described after this and at the end of this chapter the methods for the three studies will be introduced.

1.1 Theoretical Background

Good evaluation questionnaires for training should be based on theory that allows looking at processes of learning. Evaluation models can be used and the generating and selection of items should follow a certain principle of choice. Otherwise a questionnaire is nothing less than an accumulation of questions. There are different evaluation models to assess the quality and effectiveness of trainings at different levels. The article focuses on two of them, because they provide the theoretical background for the instruments having been tested.

1.1.1 The trilogy of quality from Donabedian

The first theory is the trilogy of quality from Donabedian (1966). His model consists of three consecutive levels:

- Structure,
- Process and
- Outcome.

The level quality of the structure includes all tangible and intangible resources which are required for training, for example the premises and the technique. The process level specifies how these resources are used to produce goods and services, for example if a trainer is able to perform well with a certain presentation technique. The last level in Donabedian's (1966) trilogy of quality, after structure and processes, is the outcome. This is about all the changes that take place in the individual who participates in training, for example satisfaction, attitude change or mastery of new skills. The lower level is always the prerequisite for the next one. If the structure quality of a seminar is already insufficient, it will be hard to achieve a good quality of the process and also the outcome quality will be affected. That means that the effectiveness of a training decreases with every following level.

1.1.2 Kirkpatrick's Four-Level Training Evaluation Model

Another well-known model for evaluating training courses is the one proposed by Kirkpatrick (1959, 1967 & 1976). He looks at the outcome in terms of four dimensions which are also hierarchically organized:

- Reactions,
- Learning,
- Behaviour and
- Results.

Reaction is the first level in Kirkpatrick's model (1959, 1967 & 1976), this could be satisfaction with the training for example. The second level is learning. This could be learning success, but also attitude changes. The next step would be to put what has been learned in the training session into practice. However, achieving this transfer is not a simple task. Estimates from the USA assume a transfer rate of ten percent (Baldwin & Ford, 1988). Also Kirkpatrick's last level (results) is very hard to achieve. The assumption is that the effectiveness of a training course should also have an impact on the company's key figures. However, these key figures are often also influenced by other factors. Furthermore, training sessions are usually not for everybody in the company, but just for a few. Therefore, it is very difficult to isolate the impact of one event. If these key figures could be collected, then cost-benefit analysis would be possible, as shown in the work of Sigrun Fritz (2004). Also in this model the lower level is always the prerequisite for the next one. Kirkpatrick assumes that if participants are not satisfied with a seminar, for example with a trainer, it is not very likely that they learn much or change their attitude in the desired direction, nor will they apply the learned matters.

1.2 Evaluation Instruments

The Seminar Evaluation Questionnaire (SEQ) was developed to evaluate face-to-face training by asking the participants directly after the training session. That way a high response rate can be achieved, because the target group is on site. Fresh impressions can be used to ask participants how they perceived the quality of the structure and the processes of the training. If the results indicate that something is not working the way it is supposed to, the

provider can learn from the evaluation where he can make improvements. The Transfer Evaluation Questionnaire (TEQ) was designed to measure the transfer a couple of months after a seminar and is mostly applied as an online questionnaire by using the email addresses of participants.

1.2.1 The Seminar Evaluation Questionnaire (SEQ)

Asking participants right after the training is the approach that has been used for developing the Seminar Evaluation Questionnaire (SEQ) by Masuhr et. al (2009). The underlying theory of the SEQ is the trilogy of quality from Donabedian (1966) as described above. The German questionnaire consists of 21 statements which need to be rated on a 6-point Likert scale from "strongly agree" (1) to "strongly disagree" (6). In table 1, there are the original scales and their items of the SEQ presented. No information about the verbal characterisation of the scales could be found, for example the definition of the scale "content and learning success". Also shown below are the corresponding reference values like means and standard deviations and also the sample size (N). It is possible to add questions to the SEQ. Open questions are often very useful for the trainer when receiving feedback (see table 1).

Table 1.
The original Seminar Evaluation Questionnaire (SEQ) with reference values

Item no.		N	Mean	Standard deviation
	Content and learning success			
1	I learned a lot in this seminar.	7780	1.99	.89
2	The seminar met my expectations completely.	7778	2.14	.99
3	The structure of the seminar was logical.	7773	1.99	.95
4	The contents of the seminar were communicated well.	7759	1.92	.92
5	During the seminar references to the work field were made.	7720	2.03	1.02
	Transfer motivation			
6	The implementation of the seminar contents is a very interesting challenge for me.	7692	1.97	.94
7	I got new ideas from the seminar for my occupation.	7739	2.04	.99
8	I can apply the contents practically in many cases.	7692	2.11	.97
9	I will apply the seminar contents to my work.	7660	2.00	.96
	Interaction in the seminar			
10	The participants were able to contribute their own ideas and experiences.	7731	1.77	.87
11	The participants were able to influence the seminar programme.	7710	2.25	1.07
12	The exchange of information between the participants was actively encouraged.	7725	2.06	.96
13	In the seminar, previous knowledge of the participants was acknowledged.	7654	2.21	1.02
	Seminar handouts			
14	The handouts helped to understand the seminar	7480	2.14	1.02

Item no.		N	Mean	Standard deviation
	content.			
15	I was able to work well with the handouts.	7393	2.25	1.07
16	The handouts were a good summary of the teaching contents.	7409	2.12	1.03
17	The handouts gave more depth to the subject matter.	7382	2.24	1.07
	Organization of the seminar			
18	I felt well taken care of during the seminar.	7713	1.51	.73
19	The catering during the seminar was good.	7645	1.67	.87
20	The staffs of the IAG were friendly, obliging and competent at all times.	7599	1.37	.62
21	Beforehand, I received sufficient organizational information about the seminar.	7641	1.62	.89
	Further questions			
	How did you find out about the seminar?			
	O superiors			
	O colleagues			
	O seminar programme			
	O internet presence of the IAG			
	O other (please specify):			
	Was it easy to register?			
	O yes, it was O no, it was not			
	What did you like about the seminar?			
	What could we improve?			
	What other topics should the IAG offer seminars on?			

Free translation from German into English, not tested yet

According to Masuhr et. al (2009) this is how the scales can be assigned to the levels of quality:

- Structure: organization of the seminar,
- Process: interaction in the seminar and seminar handouts, and
- Outcome: content and learning success as well as transfer motivation.

The main disadvantage of asking participants right after the training is that they cannot be asked about their behaviour after the seminar, because they have not had a chance to apply what they have learned yet. The authors point to a sub-project “training” (Gallenberger, 2007), where it was shown that transfer can be predicted by the transfer motivation of the participant at the end of a seminar. That is how they want to overcome this shortcoming, by surveying the motivation to transfer with the SEQ.

1.2.2 The Transfer Evaluation Questionnaire (TEQ)

Another approach is to question participants a couple of months after the training. That way they can provide an insight into their transfer activities. As a result, the Transfer Evaluation Questionnaire (TEQ) was designed by Masuhr based on the findings of her diploma thesis (2004). Masuhr used Kirkpatrick's model as the underlying evaluation model for this questionnaire and applied it to the field of occupational health and safety. The original scales with their items are presented in table 2. The corresponding reference values like means and standard deviations, percentages and also the sample size (N) are displayed. The German TEQ consists of 32 statements which need to be rated on a 6-point Likert scale from "does not apply at all" (1) to "fully applies" (6). Three items require a different scale. Item number 14 is: "How much could you learn in this seminar?" and the scale ranges from "very little" (1) to "very much" (6). Item number 15 is: "Which school grade would you give this seminar?" The Academic grading in Germany uses a scale from "very good" (1) to "insufficient" (6). And there is one more additional question (item 16) about the attempt to transfer ("Have you tried to apply the matters (knowledge/ procedural methods) learned during the seminar at your work place?") with a categorical answering format ("yes" and "no"). No information about the verbal characterisation of the scales could be found, for example a definition of the scale "satisfaction".

Table 2.
The original Transfer Evaluation Questionnaire (TEQ) with reference values

Item no.		N	Mean	Standard deviation
	Satisfaction			
1	Overall, I was satisfied with the seminar.	1422	5.14	.94
2	I would recommend this seminar to others.	1425	5.27	.96
3	It was worthwhile attending the seminar.	1422	5.12	1.05
15	Which school grade would you give this seminar?	1422	1.87	.72
34	The seminar was worth the expense.	1396	4.57	1.20
	Relevance			
4	The contents/topics of the seminar were relevant for my work field.	1421	4.93	1.04
5	My professional duties and problems were taken into account.	1414	4.73	1.10
6	During the seminar I already recognized where I can apply learned matters to my workplace.	1417	4.93	1.03
	Learning success			
14	How much could you learn in this seminar?	1416	4.80	1.01
23	To me personally the seminar contributed, in such a way that I learned new things by applying the seminar contents.	1419	4.94	1.05
	Assessment of the trainer *(also the handouts)*			
7	The structure of the seminar was logical.	1423	5.10	.99
8	The handouts were useful and appealing.	1419	5.09	1.01
9	The trainers had expert knowledge of the topic.	1421	5.44	.82
10	The trainers explained the contents clearly.	1418	5.31	.89

Item no.		N	Mean	Standard deviation
11	The trainers involved the participants actively.	1416	5.28	.94
12	The trainers used different learning methods (group work, role plays, videos, etc.).	1409	4.79	1.33
13	The personal contact with the trainers was very pleasant.	1416	5.47	.87
	Attempt to transfer	**N**	**Alternative answering format**	
16	Have you tried to apply the matters (knowledge/ procedural methods) learned during the seminar at your work place? (despite of whether you were successful or not)	1372	Yes = 1100 (80 %)	No = 272
	Effects on the individual *To me personally the seminar contributed, in such a way that…*	**N**	**Mean**	**Standard deviation**
17	... I now engage more in occupational safety.	1398	4.33	1.28
18	... I am a role model to others concerning OH&S.	1399	4.45	1.26
19	... I point out safety hazards more often and earlier.	1403	4.61	1.23
20	… I think more often about these topics now.	1400	4.66	1.24
21	... I realize now in which areas of OH&S I might have some knowledge gaps.	1418	4.58	1.19
22	... I see how different aspects of OH&S fit together which I haven't seen before.	1413	4.57	1.14
23	… my attitude towards OH&S has improved.	1412	4.41	1.32
	Effects on the company *To me personally the seminar contributed, in such a way that…*			
25	... unrecognized hazards have been identified now.	1389	3.53	1.50
26	… occupational health and safety has improved.	1395	3.59	1.38
27	... the company fulfils their legal obligations now.	1383	4.06	1.47
28	… employees show safer behaviour at their workplace.	1382	3.93	1.39
	Activities in the area of OH&S			
29	If colleagues ignore health and safety regulations I take action immediately.	1407	5.01	1.06
30	If I learn about safety problems, I look for solutions with others.	1408	5.25	.88
31	I am often one of the first to recognize the presence of a hazard.	1405	4.29	1.19

Item no.		N	Mean	Standard deviation
32	I set a good example when it comes to OH&S.	1408	5.04	.92
33	I can state an impact of the implemented OH&S measures.	1396	4.57	1.12

Free translation from German into English, not tested yet

This is how the scales can be assigned to the four levels of Kirkpatrick:

- Reactions: satisfaction, relevance and assessment of the trainer,
- Learning: learning success,
- Behaviour: effects on the individual and effects on the company, and
- Results: activities in the area of OH&S.

It is possible to measure transfer with this questionnaire independently from the seminar content. The disadvantage of this approach is that at this point it is too late to ask the participants about structural or procedural aspects of the training, because they might have forgotten.

2. SAMPLES AND RESPONSE RATES

The sample size for the quality test of the SEQ (1st study) amounts to 7898 people. These are all participants who attended OH&S seminars at the Institute for Work and Health (IAG) of the German Accident Insurance (DGUV) from the year 2009 to 2010. The response rate is almost one-hundred percent, because the evaluation is part of the seminar. The participants filled out the questionnaires at the end of the seminars. Usually they have 15 minutes time for that and then they get their certificates.

The dataset of the studies 2 and 3 consists of a sample from the German Social Accident Insurance Institution for the public sector. They offer about 100 seminars in the field of occupational health and safety per year with an average of 15 participants per seminar. That makes a total of about 1500 participants each year. The sample contains the ratings of 1428 people from the year 2009 to 2011. The participants filled out an online survey three months after the seminar. They received an email with a link to access the survey. After two weeks there was usually a reminder email. This reminder email has proven to be very efficient. Half the participants who participate in the survey answer after the second email. Due to this procedure, the response rate in the last three years has been between 37 and 46 %.

3. METHODS

For the quality tests in the studies 1 and 2 factor analysis, item difficulty, discriminative power and the reliability in terms of Cronbach's Alpha were calculated. Since the two measurement tools were developed based on well-researched evaluation models, the theoretical background was also tested by using bivariate correlations calculated with Spearman's rho (P). In study 3, which is based on an ex post facto design, descriptive statistics were used. The following paragraphs will explain the methods briefly.

3.1 Item difficulty index

The difficulty indices indicate the level of agreement between respondents. The value area ranges from 0 to 1. A strong acceptance or rejection of an item provides little information for evaluating seminars, therefore, the results should have an index between .2 (almost does not apply) and .8 (almost fully applies). Items with values below .2 and above .8 should be removed from the questionnaire, because they are not useful in differentiating between good and poor seminars.

3.2 Factor analysis

Factor analysis (FA) provides information about which items can be combined to form a scale. FA was performed with orthogonal rotation (Varimax). Parallel analysis was used as a criterion for factor extraction. In order to determine how many factors should be extracted, eigenvalues from the sample are compared with the eigenvalues of random numbers. The number of empirical eigenvalues which are greater than those generated randomly is also the number of factors to be extracted.

3.3 Reliability and discriminative power

To investigate the quality of the items further, selectivity was calculated. The discriminatory power of an item is the correlation of that item with the overall result of the scale. It indicates the extent to which an item represents the characteristics measured by the scale. The value area ranges from -1 to 1. A value above .7 indicates very good selectivity, between .5 and .7 is an acceptable result and below .5 is poor and these items should be eliminated. Cronbach's Alpha was calculated for the scales and for the whole questionnaire.

3.4 Investigating the theoretical framework of the questionnaires

Both evaluation models have a hierarchical structure. The lower level is always the prerequisite for the next one. For the trilogy of quality from Donabedian (1966) this means the level of structure quality is the prerequisite for the process quality and this level is the prerequisite for the outcome quality. The empirical data should show the relationships between the levels. Therefore, the correlations between adjacent levels should be higher than the correlation between levels further away from each other. The correlations which can be calculated are shown in figure 1. According to the theory "corr. a" should be greater than "corr. c" and also "corr. b" should be greater than "corr. c". Bivariate correlations were calculated with Spearman's rho (P).

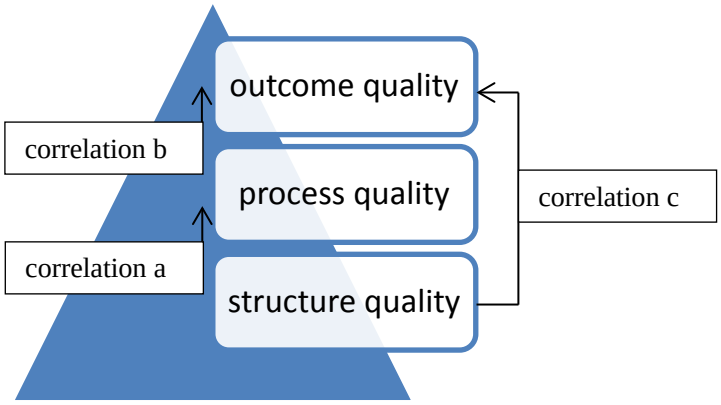

Figure 1. *Possible correlations between the levels of quality (Donabedian, 1966)*

3.5 Descriptive Statistics

In study 3 means, standard deviations and percentages were calculated to describe the results. IBM SPSS Statistics 19 was used.

4. RESULTS OF THE STUDIES

Study 1: Analysis of the Seminar Evaluation Questionnaire (SEQ)

The following paragraph will show the quality test of the SEQ and also the verification of the theoretical background. At the end of this section there will be a discussion about the results.

Item difficulty index of the SEQ

The calculated difficulty indices for the Seminar Evaluation Questionnaire ranged from .75 - .93. Nine out of 21 items exceeded the critical value of .8. All four items from the scale "organization" exceeded the critical value of .8, three from "learning success", one from "interaction" and one from "transfer motivation".

Factor analysis for the SEQ

Factor analysis corresponding to the principal component method yielded a four-factor solution, which accounted for 70.74 % of the total variance. The original version by Masuhr et al. (2009) contains five scales. The items of the scales: "content and learning success" and "transfer motivation", which are found in the original questionnaire load mostly on one factor. Also, three items of the scale "content and learning success" show loadings on two factors at the same time, so their allocation is not clear. The scales "interaction", "seminar handouts" and "organization" could be replicated.

Reliability and discriminative power of the SEQ

Selectivity of the items showed a very good result. The values ranged from .53 to .88. None of the 21 items showed a poor result. Five items were average and the other 16 showed very good discriminatory power.

Cronbach's Alpha for the scales ranged from .78 - .94 and for the whole questionnaire the internal consistency was .94.

Detailed display of the statistical values

The values of the item analyses are displayed in table 3. The results of the calculation of the item difficulty, the factor analysis and the discriminative power were integrated. This way there could be a criteria based decision about the remains of items in the questionnaire or the elimination. Exclusion criteria are usually an ambivalent factor loading or one below .5, an item difficulty index below .2 or above .8 and a discriminative power below the value of .5. In the table items are crossed out, if at least one criterion was violated.

Table 3.
Integration of the item analysis results for the SEQ

Item nr.	Scale	Item difficulty	Factor loadings				Discriminative power
			1	2	3	4	
~~1~~	~~Content and learning success~~	~~.81~~	~~.69~~				~~.742~~
2	Content and learning success	.77	.63				.798
~~3~~	~~Content and learning success~~	~~.81~~	~~.45~~	~~.53~~			~~.760~~
~~4~~	~~Content and learning success~~	~~.82~~	~~.47~~	~~.48~~			~~.789~~
~~5~~	~~Content and learning success~~	~~.79~~	~~.54~~	~~.49~~			~~.715~~
~~6~~	~~Transfer motivation~~	~~.81~~	~~.77~~				~~.784~~
7	Transfer motivation	.79	.82				.830
8	Transfer motivation	.78	.84				.853
9	Transfer motivation	.80	.82				.838
~~10~~	~~Interaction in the seminar~~	~~.85~~		~~.74~~			~~.716~~
11	Interaction in the seminar	.75		.79			.712
12	Interaction in the seminar	.79		.79			.714
13	Interaction in the seminar	.76		.70			.697
14	Seminar handouts	.77			.87		.880
15	Seminar handouts	.75			.87		.881
16	Seminar handouts	.78			.87		.883
17	Seminar handouts	.75			.77		.806
~~18~~	~~Organization of the seminar~~	~~.90~~				~~.75~~	~~.645~~
~~19~~	~~Organization of the seminar~~	~~.87~~				~~.76~~	~~.539~~
~~20~~	~~Organization of the seminar~~	~~.93~~				~~.82~~	~~.671~~
~~21~~	~~Organization of the seminar~~	~~.88~~				~~.71~~	~~.527~~

Investigating the theoretical framework of the SEQ

For the SEQ, the evaluation model from Donabedian (1966) was used. According to Masuhr et al. (2009), the scale "organization" is considered to belong to the structure. The scales "interaction" and "seminar handouts" belong to the quality of the process and "content and learning success" and "transfer motivation" are outcomes of the seminar, as described in the introduction. The model has a hierarchical structure. The level of structure quality

is the prerequisite for the process quality and this level is the prerequisite for the outcome quality. The empirical data should show the relationships between the levels. Therefore, the correlations between adjacent levels should be higher than the correlation between levels further away from each other. Bivariate correlations were calculated with Spearman's rho (P). The assumptions hold true (see figure 2). The assumption made in section 3.4 holds true since P = .666** > P = .378** and P = .402** > P = .378**. The levels of significance are p < .01 (calculated one-way) for all the correlations.

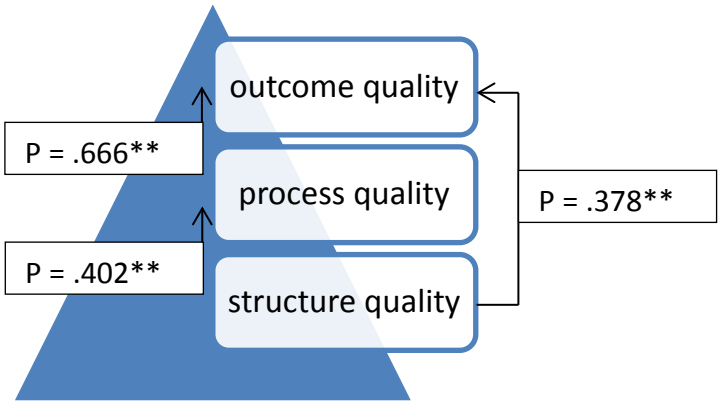

Figure 2. *The correlations between the levels of quality (Donabedian, 1966) in the SEQ*

Discussion of the analysis results for the SEQ

The quality test and the investigation of the theoretical framework show a very positive result. More than one third of the items have a very high positive acceptance. This can be viewed as being both negative and positive. Quality management of training at the IAG is very well-implemented and sets high standards; therefore, participants are satisfied very often, which is a positive result. Masuhr (2005) tried to develop a stricter version of the questionnaire, but the items still have a very high acceptance rate. It would be interesting to test the questionnaire in a company which does not have such high standards. In this situation, it is expected that values would be very different. Another argument to prove the quality of the items, despite the high positive acceptance rate, is the relationship between difficulty index and discriminate power. It is a quadratic function that can be represented as a parabola which is open at the bottom. The optimum point is where the difficulty index is .5 and the discriminate power is 1. The more the difficulty index deviates from the optimum, the lower the discriminate power is. However, the discriminate power of the items in this study is very good and so the result is acceptable and shows that the items are of high psychometric quality. Therefore the items which exceeded the critical value of .8 for the item difficulty index should not be removed from the questionnaire.

The original scales could be replicated for the most part. Due to the factor analysis the scale "content and learning success" and "transfer motivation" should be merged and could maybe be called "learning success and motivation". The items 3, 4 and 5 should be eliminated, because they also show a factor loading on a second factor. This is the same factor, where the items of the scale "interaction in the seminar" load on. By looking at the wording of item 4, it is not surprising that it has an ambivalent factor loading: „The contents of the seminar were communicated well". The fit of the 21 items on the four scales with the explained variance of 70.74 % is a good result. The only drop of bitterness is the merging of the two scales "content and learning success" and "transfer motivation". The authors wanted to overcome the shortcoming, that transfer cannot be observed with the SEQ, because it is designed to observe the target group right after training. Nevertheless the SEQ is a short questionnaire with excellent psychometric properties and a well-researched theoretical background.

Study 2: Item analysis of the Transfer Evaluation Questionnaire (TEQ)

The following paragraph will provide psychometric measures for the TEQ. Furthermore it will be proven whether the relationships formulated by the underlying theory can be confirmed with empirical data. At the end of this section there will be a discussion about the results.

Item difficulty index of the TEQ

The calculated difficulty indices for the Transfer Evaluation Questionnaire ranged from .51 - .89. Thirteen out of 33 items exceeded the critical value of .8. The items for "relevance", "learning success", "effects on the individual" and "effects on the company" show acceptable difficulty indices.

Factor analysis of the TEQ

A factor analysis corresponding to the principal component method yielded a five-factor solution which accounts for 67.43 % of the total variance. The scales “satisfaction”, “relevance” and “learning success” load mostly on one factor. Also, four items of the scale “satisfaction” scale show loadings on two factors at the same time. So does one item of the scale “learning success” and two of the “assessment of the trainer”. Therefore the allocation of the factors is not clear. The scales “effects on the individual”, “effects on the company” and “activities in the area of OH&S” could be replicated.

Reliability and discriminative power of the TEQ

Nine out of 33 items showed an average selectivity, the remaining ones showed very good discriminatory power. The spectrum ranged from .5 to .85. Cronbach’s Alpha for the scales ranged from .85 - .92 and showed a value of .94 for the whole questionnaire.

Detailed display of the statistical values

The values of the item analyses are displayed in table 4. The results of the calculation of the item difficulty, the factor analysis and the discriminative power were integrated. This way there could be a criteria based decision about the remains of items in the questionnaire or the elimination, as described in study 1. Also in table 4 items are crossed out, if at least one criterion was violated.

Table 4.
Integration of the item analysis results for the TEQ

Item no.	Scale	Item difficulty	Factor loadings					Discriminative power
			1	2	3	4	5	
~~1~~	~~Satisfaction~~	~~.83~~	~~.68~~	~~.54~~				~~.819~~
~~2~~	~~Satisfaction~~	~~.85~~	~~.65~~	~~.56~~				~~.814~~
~~3~~	~~Satisfaction~~	~~.83~~	~~.74~~					~~.850~~
~~15~~	~~Satisfaction~~	~~.83~~	~~.62~~	~~.41~~				~~.738~~
~~34~~	~~Satisfaction~~	~~.88~~	~~.58~~	~~.41~~				~~.699~~
4	Relevance	.79	.80					.706
5	Relevance	.75	.76					.714
6	Relevance	.79	.77					.748
14	Learning success	.76	.61					.711
~~23~~	~~Learning success~~	~~.79~~	~~.49~~	~~.44~~				~~.613~~
~~7~~	~~Assessment of the trainer~~	~~.82~~	~~.50~~	~~.65~~				~~.759~~
~~8~~	~~Assessment of the trainer~~	~~.82~~	~~.49~~	~~.54~~				~~.681~~
~~9~~	~~Assessment of the trainer~~	~~.89~~		~~.74~~				~~.765~~
~~10~~	~~Assessment of the trainer~~	~~.86~~		~~.82~~				~~.818~~
~~11~~	~~Assessment of the trainer~~	~~.86~~		~~.76~~				~~.778~~
12	Assessment of the trainer	.76		.62				.504
~~13~~	~~Assessment of the trainer~~	~~.89~~		~~.71~~				~~.736~~
17	Effects on the individual	.67			.81			.799
18	Effects on the individual	.69			.76			.753

Item no.	Scale	Item difficulty	Factor loadings					Discriminative power
			1	2	3	4	5	
19	Effects on the individual	.72			.81			.793
20	Effects on the individual	.73			.82			.816
21	Effects on the individual	.72			.53			.590
22	Effects on the individual	.71			.66			.735
24	Effects on the individual	.68			.72			.748
25	Effects on the company	.51				.71		.632
26	Effects on the company	.52				.81		.807
27	Effects on the company	.61				.76		.708
28	Effects on the company	.80				.80		.815
29	Activities in the area of OH&S	.80					.80	.674
~~30~~	~~Activities in the area of OH&S~~	~~.85~~					~~.82~~	~~.722~~
31	Activities in the area of OH&S	.66					.69	.583
~~32~~	~~Activities in the area of OH&S~~	~~.81~~					~~.81~~	~~.702~~
33	Activities in the area of OH&S	.71					.69	.634

Investigating the theoretical framework

The empirical data should show the relations between the evaluation levels as postulated by Kirkpatrick (1959, 1967 & 1976). Therefore, the correlations between adjacent levels should be higher than the correlation between levels further away from each other. Masuhr (2004) used the model of Kirkpatrick and applied it to the field of occupational health and safety. Therefore, the "reaction level" contains items of satisfaction, relevance and assessment of the trainer. The "learning level" is only represented by two items of the "learning success". The level "transfer" or "behaviour" contains the two scales "effects on the individual" and "effects on the company". And the last level "results" is equivalent to the scale "activities in the area of OH&S" according to the author. Again, bivariate correlations were calculated with Spearman's rho for this model. The assumptions hold true (see figure 3). The correlations of levels close to each other are bigger than the correlation of levels which are not directly next to each other. The levels of significance are $p < .01$ (calculated one-way) for all the correlations.

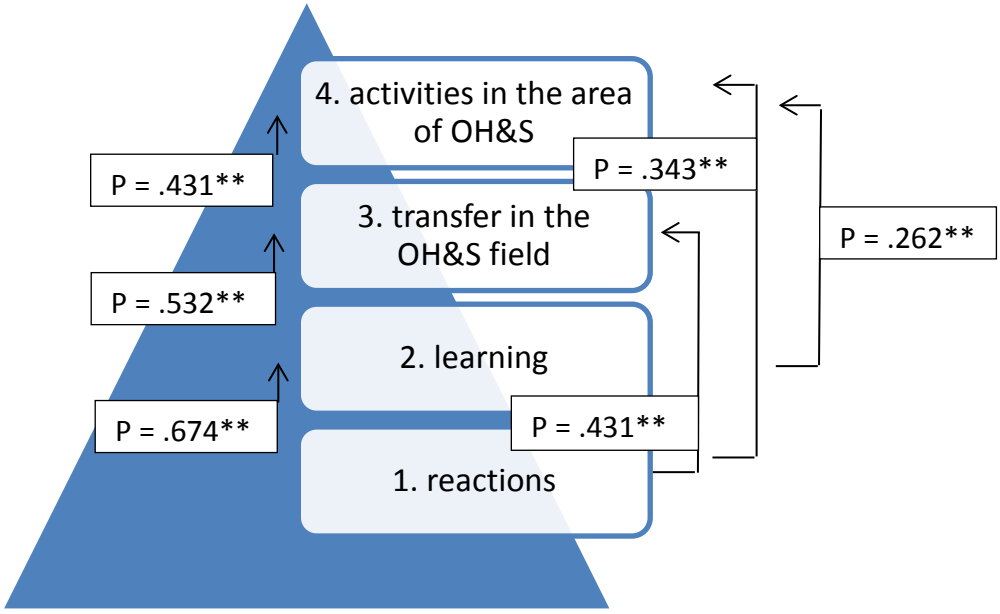

Figure 3.
Correlations between Kirkpatrick's levels applied to the field of OH&S in the TEQ

- Correlations between 1. reaction and 2. learning (P = .674**), and 2. learning with 3. transfer (P = .532**) are greater than the one between 1. reaction and 3. transfer (P = .431**)
- Correlations between 2. learning and 3. transfer (P = .532**) and 3. transfer and 4. results (P = .431**) are greater than the one between 2. learning and 4. results/activities in the area of OH&S (P = .262**)
- Correlations between all adjacent levels (P = .674**, .532**. and .431**) are also greater than the correlation between the first and last level (P = .343**)

Masuhr (2004) divided the level of transfer into two parts ("effects on the individual" and "effects on the company"), as described above. This five-level model was also tested and holds true for the assumptions as well.

Discussion of the analysis results for the TEQ

Also in this questionnaire one third of the items have a very high positive acceptance. The reasons are probably the same as those for the SEQ. The discriminate powers of the items are very good. Only nine out of 33 items showed an average selectivity. The five-scale solution explained 67.43 % of the total variance, which is a good result. The assumption of the original model of Kirkpatrick (1959, 1967 & 1976) could be confirmed by the empirical data.

The only aspect which can be seen somewhat critically is the fit of the items on the levels of Kirkpatrick. Is it really enough to ask some questions about activities in the area of OH&S and claim this would be equivalent to what Kirkpatrick described as results? By looking at the five statements on this level, it becomes clear that there is no connection to the seminar.

The statements are:

(1) If colleagues ignore health and safety regulations I take action immediately,

(2) If I learn about safety problems. I look for solutions with others,

(3) I am often one of the first to recognize the presence of a hazard,

(4) I set a good example when it comes to OH&S and

(5) I can state an impact of the implemented OH&S measures.

What is missing here is a link such as “due to the seminar”. So it is really questionable whether these five statements can operationalize the “result level”. Another point that can be viewed critically is that the level of learning only consists of two items. Factor analysis for the TEQ suggested merging satisfaction, relevance and learning success together. According to the theory, satisfaction and learning are on two different levels.

The TEQ showed good psychometric properties in general. The differentiation between different levels of transfer is an elegant solution from Masuhr (2004) and helps gain more insight into applying what has been learned. With the TEQ, it is possible to measure transfer independently from the seminar content, which is another big advantage, because it can be used for different kinds of seminars. The 33 items of the TEQ can be accompanied by further questions including some with open categories.

Study 3: Empirical data on transfer activities in the field of OH&S

The examination of the data is based on the original structure of the TEQ, therefore all the means of the TEQ scales are displayed in table 5. They show that the seminars had an effect on the participants. They are all above the value of 3.5, which is the middle of the 6-point Likert scale (ranging from one to six). The satisfaction with the seminar was rated with a 4.6 on average and relevance and learning success with a 4.9. The one mean that stands out is the very strong rating of 5.3 for the assessment of the trainer. At the level of individual effects the participants were asked to state

(1) whether they now engaged more in occupational safety,
(2) if their attitude towards it has improved,
(3) if they think more often about these topics now,
(4) if they are a role model to others concerning OH&S,
(5) if they point out safety hazards more often and earlier,
(6) if they realize now that they might have some knowledge gaps and
(7) if they see how different aspects of OH&S fit together which they hadn’t seen before.

They rated this with a 4.5 on average. On the company level the participants were asked to rate four statements, as already described:

(1) whether unrecognized hazards have been identified now,
(2) whether occupational health and safety has improved,
(3) whether the company fulfils their legal obligations now and
(4) whether the employees show safe behaviour at their workplace.

The participants rated that these four statements on average “quite apply” (3.8). The rating for the last level (activities in the area of OH&S) is quite high with 4.8.

Table 5.
The results of the original TEQ scales

TEQ scales	N	Mean	Standard deviation
Satisfaction	1328	4.57	.59
Relevance	1404	4.87	.93
Learning success	1408	4.87	.94
Assessment of the trainer	1298	5.26	.79
Effects on the individual	1387	4.52	1.01
Effects on the company	1364	3.78	1.23
Activities in the area of OH&S	1384	4.83	.81

Participants were also asked, whether they tried to apply the matters learned during the seminar yet and 80 % of the participants claimed that they did. The result can be seen in table 6.

Table 6.
Result of the item "attempt to transfer"

Attempt to transfer	N	yes	no
Have you tried to apply the matters (knowledge/ procedural methods) learned during the seminar at your work place? (despite of whether you were successful or not)	1372	1100 (80 %)	272 (20 %)

The theoretical background of the TEQ is the evaluation model of Kirkpatrick, therefore the means of the levels were also observed (see table 7). It is visible, that they also exceed the value of 3.5, hence a change in the participants can be stated. What stands out is the result of the last level with the value of 4.8. The mean appears very high, by looking at the tendency of decreasing values from level 1 to level 3. But the mean of level 4 does not fit in this tendency.

Table 7.
Results of the TEQ on the four levels of the theoretical framework

Levels	N	Mean	Standard deviation
1. Reaction	1290	4.91	.68
2. Learning	1408	4.87	.94
3. Transfer in the OH&S field	1345	4.15	1.01
4. Activities in the area of OH&S	1384	4.83	.81

Discussion on the empirical data on transfer activities in the field of OH&S

The empirical data shows that the observed seminars of the German Social Accident Insurance Institution for the public sector were rated very positive. The result that 80% tried to put into practice what they have learned in the seminar is great. However, trying is not the same as being successful. Thus, the interesting question is what happens after someone fails. Do they try again and find different ways? Did the seminar maybe already prepare them for obstacles and how to overcome them? It would be interesting to have a couple open questions accompanying the items of the TEQ. The high means of the scales show that transfer takes place and even leads to some changes at the company level, because the value here is also above 3.5.

One big weakness of this study is the design. The study has an ex post facto design, which means there was no testing beforehand and there was also no control group to compare the results to. Especially for the items of the scale "activities in the area of OH&S" a "before and after" comparison would be interesting. Due to the results it is not clear, whether the high value of 4.8 has anything to do with the seminar. As already described in study 2 the link to the seminar is missing in the wording of the items. Another hint that the items from the level "results" are poorly formulated is the high value of this mean in table 7. Due to the theory, the means should be decreasing by going up the levels from one to four. The mean of the fourth level is therefore too high.

5. CONCLUSION AND OUTLOOK

The results of this article show that it is possible to evaluate seminars and training at other levels than just on satisfaction or other reactions. The SEQ and the TEQ are of good psychometric quality. Some small adjustments would be needed on both measures according to the quality tests, especially by looking of the results of the factor analysis. Because of the high acceptance rate, it would be interesting to use both questionnaires in

fields where a quality management of training is just starting off in order to get more variance in the results. This could answer the question whether the items are of poor item difficulty or the observed seminars were simply of great quality.

It has to be added at this point that more measurement is needed for both questionnaires. Measuring the reliability in order to decide whether the questionnaires measure accurately would be important. An examination of their validity is also missing. Do the questionnaires really measure what they claim to measure? What is needed is an external criterion. For example, transfer could also be measured with a second objective method such as observation by co-workers or superiors. These measurements could then be compared with the self-assessments.

User of the TEQ should be careful with the scale "activities in the area of OH&S". It is doubtful that those items really represent the "result level" according to Kirkpatrick's model (1959, 1967 & 1976). As already stated in the introduction part, it is challenging to measure this level. With the TEQ it could be shown that, in this one sample, transfer takes place and even leads to some changes at the company level. As already mentioned, it was an ex post facto design. Studies with stronger designs should be used in the future. What is also still missing is the link between learning success and transfer motivation and the actual transfer. By using the SEQ and the TEQ together it should be possible to prove whether these are predictors for transfer.

By looking at both questionnaires and their theoretical framework, it could be argued that there is a difference between the effectiveness and the quality of training. Masuhr and her colleagues (2009) defined quality of OH&S seminars as "the degree to which a set of inherent characteristics fulfil the requirements" (DIN EN ISO 9000:2000). This understanding aligns with the achievement of objectives. Proving this is one of the core functions of evaluation research. Therefore, an effective seminar is ultimately of good quality. Models like the one from Kirkpatrick allow some aspects to be looked at in more detail and are designed for training programmes. Therefore, it is strongly recommended to use evaluation models where the output can be split up and behaviour after training can be observed.

6. REFERENCES

Baldwin, T. & Ford, J. K. (1988). Transfer of training. Personnel Psychology, 41, 63-105.

DIN EN ISO 9000: 2000

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

Fritz, S.: Mehrebenen-Evaluation von Maßnahmen der betrieblichen Gesundheitsförderung. Dissertation an der Technischen Universität Dresden 2004.

Gallenberger,W. (2007). Qualität in der Prävention – Teilprojekt Qualifizierung. http://www.dguv.de/iag/de/forschung/forschungsprojekte_archiv/qdp/qdp_abschluss/_dokumente/qdp_ab10.pdf

Kirkpatrick, D. L. (1959). Techniques for Evaluating Training Problems. Journal of the American society of Training Directors, 3-26.

Kirkpatrick, D. L. (1967). Evaluation of training. In R. L. Craig & L. R. Bittel (Eds.), Training and development Handbook. New York: AS TD/Mc Graw Hill.

Kirkpatrick, D. L. (1976). Evaluation of Training. In R. L. Craig & L. R. Bittel (Eds.), Training and Development Handbook. New York: AS TD/Mc Graw Hill.

Masuhr, Kati: Transferevaluation in der Bildungsarbeit: Entwicklung eines Instruments zur Erfassung von Transfer in berufsgenossenschaftlicher Aus- und Weiterbildung. Dresden, Technische Universität, Diplomarbeit, 2004

Masuhr, K., Windemuth, D. & Taşkan-Karamuersel, E. (2009). Development of an evaluation instrument to predict effectiveness from training in occupational health and safety. Safety Science Monitor, 13, 2, 1-8. http://ssmon.chb.kth.se/vol13/issue2/index.php