MASTER THESIS

A Case Study on Exploring the Relevance of Evaluation Characteristics in Designing an Evaluation Approach on Behaviour Level on Training Effectiveness

Researcher	**Soraya Mohammad**
Student number	**s1623575**
Faculty	**Behavioural Sciences**
Educational program	**Master Educational Science & Technology**
Educational Institute	**University of Twente**
Commissioned by	**Deloitte Nederland**
Graduation Committee	**B.J. Kollöffel**
	M.D. Endedijk
External supervisor	**J. Ketting-Antonis**
Date	**July 8th 2016**

UNIVERSITY OF TWENTE. **Deloitte.**

Acknowledgements

Summary

The current business environment changes due to technology, business trends and organisations striving for a competitive advantage. As a result, more organisations focus on learning opportunities and training for their professionals and evaluate these programs to determine if the current way of working is, or if elements can be adapted or changed. Evaluation can be done on several levels: reaction, learning, behaviour and results. The two levels of evaluation that are mostly related to transfer and effectiveness of training, are *behaviour* and *results.*

This present study describes an exploratory case study, which was set out for Deloitte Netherlands, a large organisation, that employs professionals to provide audit, tax, consulting, financial advisory and risk management to clients. The aim of this study was to investigate what evaluation characteristics are seen as relevant by learning professionals in designing an evaluation approach on behaviour level with the goal to determine the effectiveness of the training. By going back to the literature and the evaluation frameworks that are used to guide an evaluation, evaluation characteristics are identified and there is explored what these characteristics mean for the theory and existing evaluation frameworks. Three semi-structured focus-groups interviews with the learning department of Deloitte provided input on identifying evaluation characteristics and the role of them in designing an evaluation approach. There is taken into account what types of training allow an evaluation on behaviour level, the possible measurement methods, and involvement. Next to that, input was given on conditions that need to be met to evaluate, time and frequency for evaluation, and possible barriers or challenges.

Results showed that skill trainings, contrary to knowledge training, seem suitable for an evaluation on behaviour level according to the learning professionals. To measure the behaviour, it is found that a performance manager of a trainee could play a key role in the collection of data by observing behaviour on the job. For this, measurable behaviour indicators could be used. To ensure validity and reliability, organisations may include a peer report or feedback from a subordinate. A condition that seemed important, is a firm collaboration between the business units within the work field and the training department, to enrich the relationship. A suggestion is made regarding the moment of evaluation: this could be done 3 months after the training and the second one after 6 months. An evaluation could be done twice a year. The main barriers and challenges that may arise in conducting an evaluation, are difficulty in accessing data, time, and the relevance of using trainee behaviour as a predictor instead of trainee reaction. Results and literature showed that these evaluation characteristics are related to the transfer performance and therefore this study proposed to integrate evaluation characteristics into the IMTEE and recommended further research on defining the relationship between evaluation characteristics and transfer performance.

Table of Contents

1   Introduction

The current business environment changes due to technology, business trends and organisations striving for a competitive advantage (Santos-Vijande, López-Sánchez, & Trespalacios, 2011). Organisations have an increasing drive to learn, reskill and change rapidly to keep competing with their competitors and to stay relevant (Bersin, 2015). As a result, more organisations focus on learning opportunities and training for their professionals. By offering training, professionals upgrade their skills and knowledge. This contributes in meeting the business objectives (Crocetti, 2002). When these business objectives and trainings are aligned, learning and development could be used as a strategic function to improve business results and maintain a competitive advantage (Wick, Pollock, & Jefferson, 2010). When organisations focus on reskilling and training, the term Organisational Learning (OL) is used (Santos-Vijande et al., 2011).

If organizations wish to benefit from their investment in training and improve business results, a transfer of training is desired. This means that professionals apply their trained skills, knowledge and attitudes in the workplace and on-the-job (Saks & Burke, 2012). Organizations may measure a transfer of learning by evaluating (Saks & Burke, 2012). An evaluation provides data about training results and information about the extent to which training programs have achieved the determined objectives (Bates, 2004). This data is useful to determine if the current way of working is, or if elements can be adapted or changed. Evaluating provides data for decision-making based on evidence and data. This type of decision-making is called data-driven decision making or data-based decision making. According to literature, there are several levels on which can be evaluated (e.g. Kirkpatrick, 1994; Kraiger, 2002; Alvarez, Salas & Garafano, 2004). The two levels of evaluation that are mostly related to transfer and effectiveness of training, are trainee behaviour and business results (Saks & Burke, 2012).

Nevertheless, many organisations do not evaluate trainee behaviour or business results, but only evaluate on the primary reaction provided by the trainee, and do not regularly measure the impact of training on job performance and behaviour (Kraiger, 2002). Trainee reaction is the rating and appreciation that a trainee gives after the training by filling out an evaluation form and this is not considered an outcome of training. Alvarez et al. (2004) suggest that evaluating on only trainee reactions is useful to determine if the training content and design are appropriate, but does not say something about the effectiveness of training. Therefore, evaluation models suggest to evaluate on behaviour level or business results.

However, there does not exist an evaluation approach that provides relevant and useful data that says something about the effectiveness of a training. Evaluation tools are used for the primary reaction of the trainee, but there does not exist a fixed instrument or policy on measuring the effectiveness of a training. Decisions about the continuation or adaption of a training program within organisations are currently done based on the amount of cancellations, the evaluation of the trainee right after the training, the opinion of the facilitator about the training and budgetary information. According to Spitzer (2005) this is an outdated way of evaluating and today's focus is more on measuring effectiveness and return on investment (ROI).

When organisations evaluate trainee behaviour, there could be determined if the trained knowledge, skills and attitudes are transferred to the workplace and that the business objectives are achieved. By demonstrating that training is effective for the improvement of business results, the impact of learning will be shown and investments are justified (Spitzer, 2005). Management often scrutinizes the budget that is available for training. Learning might be seen as less important than the core business activities (Pulichino, 2007). Internal marketing and communication may be necessary to promote the opportunities of training to the professionals, but also to convince the management about the importance of training (Crocetti, 2002). By convincing the management that investing in learning leads to a greater organisational success, upper management might consider a larger investment (Wick et al., 2010). This success can only occur when training programs are evaluated and when its effects and the investment can be justified and the effect can be demonstrated.

The demonstration of effectiveness by evaluating the behaviour change in practice is something that is new for most organisations. Therefore organisations might adjust their current way of working concerning the evaluation on the effectiveness of training programs (Aggestam, 2006). However, learning departments who are responsible for the evaluation face many options about when to evaluate, what information might be relevant, what stakeholders should be consulted, what methods might be used, what should be measured and how they can facilitate the evaluation (Hashim, 2001). Next to that, literature provides a wide range of approaches and scientific methodologies which most often focuses on factors within the training context (Tannenbaum et al., 1993). As a result, learning departments of organisations face difficulties in choosing what approach is suitable for their particular organisation, because the characteristics in a work environment are often different in every organization (Hashim, 2001). According to Tannenbaum et al. (1993) the work environment influences the transfer of training and the behaviour that is shown on the job and this topic is significantly less researched. Therefore they recommended for future research to incorporate literature on organisational learning and their corresponding characteristics. Alvarez et al. (2004) confirm this almost ten years later, that the field of organisational characteristics did not receive much attention. To the knowledge of this research this is still an area that has not been researched a lot. As stressed by Tannenbaum et al. (1993), a learning organization has multiple characteristics that seem to be receptive for learning on an organisational level. Next to that, not evaluation characteristics are important as well. Therefore, it is important that existing frameworks or evaluation models should be complemented with evaluation characteristics of learning organizations. The aim of this study was to investigate what evaluation characteristics are seen as relevant by learning professionals in designing an evaluation approach and what this implicates for the current theory and evaluation frameworks. Therefore a literature study is conducted which investigated what organisational learning is and what evaluation characteristics, with a focus on behaviour level, are already known in literature. In a case study the perception of learning professionals from a multinational in professional services is used to determine what evaluation characteristics seem relevant in designing an evaluation approach on behaviour level.

2   Theoretical Framework

2.1   Organisational learning and training

Organisations have an increasing drive to learn and change rapidly to keep competing with their competitors (Santos-Vijande, et al., 2011; Bersin, 2015). As a result, more organisations focus on learning opportunities for their professionals. These learning opportunities are a part of OL. Within this study OL is defined as ''the transfer of knowledge that happens through dynamic two-way processes between individuals, groups and the organisation which provides new insights and understanding to improve performance'' (e.g. Aggestam, 2006; Santos-Vijande et al., 2011).

Providing training is one of the ways for organisations to improve performance. Training is seen as ''communication directed at a defined population for the purpose of developing skills, modifying behaviour, and increasing competence'' (Alipour & Salehi, 2009, p. 63). In literature, learning is referred to as ''to gain knowledge, skill, ability'' (Kind, 1968, p. 125). To understand this definition, the terms knowledge, skills and abilities will be clarified. Knowledge is defined by Huber (1991) as'' the interpretations of information, beliefs about cause-effect relationships or '' know-how'' (p.89). Skills are defined as the required capacities for performing a task and doing, and abilities are defined as general capacities related to the performance of tasks that have been developed over a period of time resulting from experience (Alipour & Salehi, 2009). AlYahya and Norsiah (2013) distinguished the skills into two categories: skill improvement and behavioural changes.

The role of training in organisations is seen as making sure that employees are enabled to perform on a high level. Organisations offer different training methods, but a distinction can be made between off-the-job training and on-the-job training. Alipour and Salehi (2009) state that off-the-job training is often located outside the workplace and is mostly organised in a group. Examples can be

lectures or simulations. On-the-job training is mostly organized while working and are directly related to job requirements. For most companies, a lot of larger training programs are organized in a formal setting and off-the job, but there also e-learnings and on-the-job learning. However, for this research the focus will be on the training programs that are offered off-the-job and in a formal context.

2.2   Training effectiveness and transfer of training

Effective training programs contribute to the organisation and performance (Spitzer, 2005; Alipour & Salehi, 2009). A training should be used to enhance business performance in order to be effective. But often the question is raised: when is a training program effective? Training effectiveness is interpreted in many ways, which makes it difficult to measure. Training effectiveness is conceptualized by Tracy, Hinkin, Tannenbaum and Mathieu (2001) as a combination of training efforts and transfer of training. Broad and Newstrom (1992) point this out as well. However, training efforts will not have the desired effects if the trained skills, knowledge and behaviours are not applied in the workplace (Aguinis & Kraiger, 2009). When the trained skills, knowledge and attitudes are applied on the job and trainee behaviour changes as a result of training, this is called 'transfer of training' (Baldwin & Ford, 1988; Seyler, Holton, Bates Burnett, & Carvalho, 1998; Rajee, Madan, & Jayarajam, 2009; Blume, Ford, Baldwin, & Huang, 2010; Grossman & Salas, 2011; Saks and Burke 2012; Weisweiller, Nikitopoulos, Netzel, & Frey, 2013). Within this research the statement of Baldwin and Ford (1988) is used: transfer of training and thus the application of the trained skills, knowledge, and changing behaviour on the job, is an indicator for training effectiveness.

2.3   Training evaluation

In order to determine the effectiveness of training, organisations could evaluate their training programs. An evaluation provides evidence that supports the decision-making process about the continuation or adaption of a training (Wick et al., 2010). Next to that, it is 'a systematic process of collecting data in an effort to determine the effectiveness and/or efficiency of training programs and to make decisions about training' (Saks & Burke, 2010, p. 119).

    As mentioned in the introduction, many organisations evaluate their training programmes by using a 'happy sheet'. This way of evaluating is measuring the reaction of the trainee by a questionnaire at the end of a training programme (Smith, 1990). A deeper evaluation might provide more insight on the effectiveness of training. Hashim (2001) admits that evaluation does not completely determines the effectiveness of a training program, because conditions and locations and settings might change. However, evaluating does provide support and an estimation of the effectiveness of training programs can be given.

As discussed in the introduction, several evaluation models and frameworks are known. The most used evaluation model is Kirkpatrick's four-level approach (Kirkpatrick, 1994; Aguinis & Kraiger, 2009; Saks &Burke, 2010). This model has four levels that are used to evaluate: reactions level, learning level, behaviour level and results level. Many models uses the Kirkpatrick's model as a basis and expanded the model, because several studies criticized the model on not taking into account intervening variables that may affect learning, program elements that are related to the context and the use of effective resources (Frye & Hemmer, 2012; Bates, 2004). Despite the criticism, Kirkpatrick's four-level approach is seen a useful taxonomy (Frye & Hemmer, 2012). However, when these levels are integrated in a more elaborate model that includes influencing variables and relationships, the four levels offer a more useful way to define effectiveness.

    A model that has tried to elaborate and integrate literature, is the Integrated Model of Training Evaluation and Effectiveness (IMTEE) that is presented by Alvarez, Salas, and Garofano (2004) (figure 1). This model builds on the most common evaluation theories of Kirkpatrick (1976), Kraiger (2002), Holton (1996) and Tannenbaum et al. (1993).Within this model there four levels of Kirkpatrick's are used. In the second model, Tannenbaum et al. (1993) elaborated Kirkpatrick's model by adding posttraining attitudes and dividing behaviour into two outcomes for evaluation: training performance

and transfer performance. Holton expanded this model by including learning, transfer and results and left out the reaction level. Finally, Kraiger's target areas of the decision-based evaluation model are used (Training content & design, changes in learners and organisational payoffs),

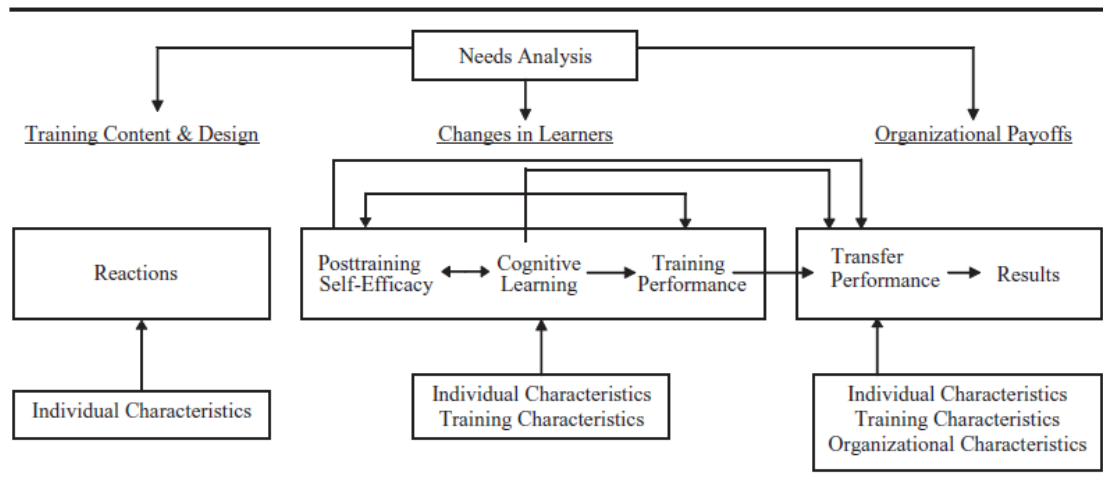*Integrated Model of Training Evaluation and Effectiveness*

Figure 1 The IMTEE model. Adapted from '' An Integrated Model of Training Evaluation and Effectiveness'', by   Alvarez, Salas, & Garofano, 2004, *Human Resource Development Review,* p.393

The model starts with a needs analysis, which is seen as the first level. This needs analysis is used to develop three main areas: the content of the training, which should enhance a change in the learner, resulting in an organisational payoff (level 2). The three main areas of evaluation are derived from Kraiger's model (2002) and are the basis for the six evaluation measures (level 3) that Alvarez et al. created by combining the theory of Kirkpatrick (1976), Holton (1996), and Tannenbaum et al. (1993. From left to right, this means that training content and design can be evaluated by the measurement of trainee reactions. Changes in learners can be evaluated by the measurement off posttraining self-efficacy, cognitive learning, and training performance. The organisational payoffs can be evaluated by the measurement of transfer performance and results.

The literature review by Alvarez et al. (2004) showed that there is a relationship between the evaluation measures, but not a hierarchical relationship. This means that the measurement levels were not seen as predictors for the next. Next to that, a training program cannot be evaluated when it is isolated from a surrounded context. Without effectiveness variables, there cannot be explained if the outcomes are because of the training design, the trainee or the organisational context. Therefore the IMTEE model includes individual characteristics, training characteristics and organisational characteristics (level 4). Individual characteristics have been researched a lot and are difficult to specify. However, the main characteristic within this category is motivation. There is also research conducted on training characteristics (e.g. instructional techniques and learning principles). The least researched characteristics are the organisational characteristics.

Reactions
 The first evaluation measure is reactions, which measures the reaction of the trainee on a training program and is connected to the training content and design. The IMTEE model suggested, in line with the theory of Alliger et al (1007), Kirkpatrick, and Kraiger, that reactions could be used to evaluate the training content and design, but a relationship with the other measures was not found. Therefore this evaluation measure is not connected with any other measurement level in this model. However, there are influencing variables within this model that are connected to the reaction level: individual characteristics. Alvarez et al. tried to identify specific variables within the model and found that two

individual characteristics, pretraining self-efficacy and motivation, seemed to influence the employee's reaction.

Posttraining self-efficacy
Posttraining self-efficacy is seen as trainees' confidence or capabilities to apply newly learned knowledge and skills in the work environment after a training (Gist, Stevens & Bavetta, 1991). The IMTEE model includes several posttraining attitudes, but found that only posttraining self-efficacy has a positive relationship with cognitive learning, training performance and transfer performance (Alvarez et al., 2004).Alvarez et al. did not find a direction of a relationship and therefore this the relationship might be reciprocal. Alvarezt et al. (2004) found in their study that posttraining self-efficacy is related to individual characteristics (e.g. pre-training self-efficacy and experience). The influencing variables that are applicable to the posttraining self-efficacy are not only individual, but also training characteristics are related to posttraining self-efficacy (e.g. learning principles and maintenance interventions). Organisational characteristics are not related to posttraining self-efficacy, therefore they are also not included in the model.

Cognitive learning
Cognitive learning is described by Alvarez et al. (2004) as ''the cognitive acquisition of knowledge and is typically measures through paper-and-pencil or electronically administered test of information taught in training'' (p.396). Cognitive learning has a reciprocal relationship with posttraining self-efficacy and is related to training performance and transfer performance. The influencing variables of cognitive learning are individual characteristics and training characteristics, and again organisational characteristics are not related to cognitive learning,

Training performance
Training performance is seen as being able to use a new skill after a training, before the transfer has occurred, and can be measured by a trainee demonstration. Kraiger (2002) stated that training performance can have two dimensions. The first dimension is being able to replicate the learned behaviours and the second dimension is performing the newly acquired skill fluidly short after training. This means that when trainees showed their skills right after training, it is still possible that the transfer to the work environment is not made. However, when comparing the relationship with other evaluation measures, Alvarez et al. (2004) indicated that training performance is influenced by posttraining self-efficacy and cognitive learning. Next to that, it has a positive influence on posttraining self-efficacy and transfer performance, as can be seen in figure 1. As was the case in the previous evaluation measures, the influencing variables that are related to training performance are individual characteristics and training characteristics.

Transfer performance
Transfer performance is explained as 'behavioural changes on the job as a result of training and can be assessed via supervisor evaluation of on-the-job behaviour or posttraining retest several months after training using the same or an alternate form of the training performance test'' (p. 397). The arrow from transfer performance to results in figure 1 suggests that there is a positive relationship between transfer performance and results. This evaluation measure can be used to say something about the organisational payoffs. The influencing variables that are related to transfer performance, are on individual level, training level, but also organisational characteristics are related. Because this study focuses especially on transfer performance in organizations, the organisational characteristics will be explained more thoroughly. A positive transfer environment was according to Alvarez et al. (2004) the only characteristic that was related to transfer performance. According to Rynes and Rosen (1995), there are 4 elements in the transfer environment that are related: perceived supervisor support, mandatory attendance for managers, rewards for practicing skills, and follow-up evaluations that are not only based on reactions. The organisational characteristics organisational commitment and posttraining mastery

orientation were not included in the model, because limited studies found a positive relationship between these variables and transfer performance.

Results
The last evaluation measure is results. This measure is described by Alvarez et al. (2004) as organisational changes, which can be related to trainee's behavioural changes as a result of training. There is not enough research done on organisational characteristics according to Alvarez et al. (2004) and other effectiveness variable in order to formulate key characteristics. They state that there might be other effectiveness variables, depending on the environment and training program and therefore this model does not completely unfold the field of training effectiveness.

Conclusion
The IMTEE is the first model that investigated relationships between the evaluation measures and influencing variables and integrated literature from 1993 to 2004 to provide a good overview of research on training effectiveness. The IMTEE added a fixed set of criteria and only used effectiveness variables that are consistently found as influencing in the training outcomes. However, the model does not include what type of evaluation should be used to determine if there is a transfer performance, indicating the effectiveness of training. These so called evaluation characteristics are essential for designing an evaluation approach that suits an organisation. According to Sachdeva (2014) there has been done a lot of research on training evaluation and effectiveness, but less research is done on evaluation of behavioural training. The reason for this limited research mentioned by Sachdeva (2014) is that there are methodological limitations. However, Sachdeva states that based on review and several analyses of studies, there is a strong need for more research in the field of evaluation of behavioural training programs to find an effective way to evaluate these programmes. Kraiger et al. (1993) already mentioned more than ten years ago that the models on training evaluation and effectiveness lack to include sound measures of learning, due to incomplete theories of what learning is and how it should be evaluated. Therefore they have suggested a nomological network in which concepts, measures and interrelationships are specified. The IMTEE model describes the concepts, influencing factors and interrelationships, but the measures and specific evaluation characteristics are left out. Therefore this research aimed to describe what evaluation characteristics are relevant according to learning professionals in choosing an evaluation approach in determining the effectiveness of the offered training programs. Next to that, there is aimed to elaborate the IMTEE with these evaluation characteristics.

2.4   Evaluation characteristics
In order to define what evaluation characteristics are relevant in designing an evaluation approach on behaviour level, this research tried to include the basic criteria for an evaluation. Therefore the pragmatic view of Hashim (2001) is used, who claimed that for the evaluation many professionals face various options about when to evaluate, what information might be relevant, what stakeholders should be consulted, what methods might be used, what should be measured and how they can facilitate the evaluation. These factors are depended of organisational and situational characteristics. The basic criteria that are chosen for this research are:
- Measurement methods
- Involvement
- Conditions
- Barriers and challenges
- Time and frequency

Therefore the following paragraphs explain the various basic criteria evaluation characteristics that might influence training outcomes, transfer of training and a pragmatic paradigm for the identification of these characteristics is used. This research took into account aspects that are related to the evaluation process and depend on the organisational context: measurement methods for behaviour, involvement in evaluation, barriers and challenges.

2.4.1 Measurement methods

A behaviour change is not something that can be measured easily. AlYahya and Norsiah (2013) described that measuring behaviour can be done in several ways: surveys, observations and interviews of performance are seen as methods to measure a change in behaviour. However, a behavioural change requires some time for the employee to have implemented the acquisition of knowledge and skills. Behavioural change may appear in attitude or communication to a supervisor or peer (AlYahya & Norsiah, 2013).

In literature several methods for the measurement of behaviour are proposed. First, Hashim (2007) investigated several measurement methods that are frequently used in organisations to evaluate training programs. Table 1 provided insight in the most common types of evaluation methods and how respondents frequent the methods are used. As can be seen in the figure, trainee feedback has the highest mean and is most frequently used. This is followed by observation, performance analysis and reaction form. An organisational audit and a peer report are used the least. Because not every evaluation measure that is found in literature can measure a behaviour change, a selection is made in this research. The methods that Hashim (2007) mentioned, that are suitable for measuring a behaviour change in the eye of the researcher are an observation, interview, peer report, supervisor, performance analysis, site visit and self-report. These methods might allow the detection of a behaviour change as a result of training.

Table I

Distribution of respondents by evaluation method

Methods	Percentage of responses					Mean
	1	2	3	4	5	
Survey	6.4	10.6	34.6	26.6	22.3	3.48
Interview	3.2	8.5	26.6	43.6	18.1	3.65
Observation	1.1	7.4	21.3	46.8	23.4	3.84
Document review	4.3	18.1	31.9	30.9	14.9	3.34
Organisational audit	22.3	20.2	31.9	19.1	6.4	2.67
Performance analysis	6.4	10.6	29.8	26.6	26.6	3.56
Expert review	19.1	17.0	37.2	19.1	7.4	2.79
Panel checklist	28.7	14.9	24.5	23.4	8.5	2.68
Site visit	11.7	21.3	26.6	29.8	10.6	3.06
Pilot tests	28.7	11.7	25.5	26.6	7.4	2.72
Trainee feedback	5.3	1.1	12.8	30.9	50.0	4.19
Simulation	23.4	22.3	26.6	19.1	8.5	2.67
Self-report	10.6	22.3	35.1	21.3	10.6	2.99
Work samples	18.1	19.1	25.5	27.7	9.6	2.91
Peer report	14.9	26.6	35.1	18.1	5.3	2.72
Supervisor	12.8	14.9	34.0	27.7	10.6	3.08
Competency test	18.1	12.8	28.7	23.4	17.0	3.08
Reaction form	11.7	8.5	28.7	21.3	29.8	3.49

Notes: 1 = not at all, 2 = seldom, 3 = sometimes, 4 = frequently, 5 = very frequently; $N = 94$

Table 1 *Evaluation methods. Adapted from '' Training evaluation: client's roles'', by J. Hashim, 2007,* Journal of European Industrial Training*, p. 376.*

Second, AlYahya and Norsiah (2013) mentioned that the use of a survey, observations and interviews of performance are used tools in measuring behaviour. There is believed that by linking the performance interviews, changes could be recognized by reflecting.. A method that is often used to measure a behaviour change, is to develop performance goals in an earlier stage. When these goals are accomplished, there is proof that the learned knowledge is translated into behaviour (AlYahya & Norsiah, 2013).

A third possibility to measure behaviour is to translate the desired behaviour into observable behaviour indicators, preferably quantitative measures (Gielen & Indesteege, 2004). By making behaviour concrete, relevant and reliable conclusions can be made. Because observable behaviour indicators are based on an opinion, a measurement scale could be used, as proposed by Gielen and Indesteege (2004.

    An element that is mentioned in literature, is that when decision makers wish to increase the amount of confidence and wonder if the program was effective and rule out external influencing factors, a plausibility assessment can be done (Habicht et al., 1999). This allows an evaluator to control for influencing factors by using a control group. A way to do this is to make a historical control group and compare behaviour for and after a program. Another option is to make an internal control group and compare the behaviour between groups within an organisation, as mentioned by Habicht et al. (1999).

    In conclusion, there can be stated that literature defined several possible measurement methods to evaluate behaviour. Using trainee feedback, an observation, performance analysis or a survey are the main methods that literature defined as suitable. A control group can be used to increase the amount of confidence.

2.4.2   Involvement

Within organisations often the practical question is raised who should be responsible and involved in the evaluation and measurement process. There should be discussed who will be responsible for collecting and processing data, but there should also be determined what the source of data will be. For the source of data, it is assumed in literature that evaluations on behaviour changes are more successful when it is measured by sources than by the trainee self (Tracey et al., 1995; Lance, Kavanagh & Brink., 2002). The sources that are mentioned are a supervisor or manager, peers and clients. These sources seem less subjective than the trainee self, however the collection of information of other sources than the trainee is more challenging and time-consuming (Velada, Caetano, Michel, Lyons, & Kavanagh, 2007). There are methods that prefer to collect the evaluation data from the trainee self. Self-reports based on behaviour are being criticized, but are also seen as an important source of data to Velada et al. (2007). Data that is collected from other sources might give other ratings than a trainee's self-perception. However, Velada et al. (2007) concluded that a trainee self-report is able to give the same results as compared to external sources. Another element that is mentioned why self-reporting of a trainee is beneficial for organizations, is that it is less expensive for an organization, because other sources can spend their time on different activities.

2.4.3   Conditions

The term conditions within this research is used to stress that in order to create and conduct an evaluation in an organisation, organisations might first determine what conditions should be met in

order to conduct a proper evaluation. These conditions differ in every organisation. For one organisation it can be budget or time, for other organisations this might be flexibility of the employees.

2.4.4 Barriers and challenges

Pulichino (2007) investigated why organisations do not regularly measure trainee behaviour and results. Results of his research showed that training professionals are not able to identify the appropriate metrics, or retrieve the data, and have difficulties to link results to training. Figure 2 shows the most important perceived barriers for level 3 Evaluation (trainee behaviour). One of the barriers is that there is no management support. According to Wick et al. (2010) it is important that the management supports the trainee in the learning experience and in the evaluation process and according to Pulichino's survey that is one of the barriers for not evaluating on trainee behaviour and business results. In designing an evaluation approach, it is recommended by Pulichino (2007) to include the barriers and challenges in order to design a fitting approach.
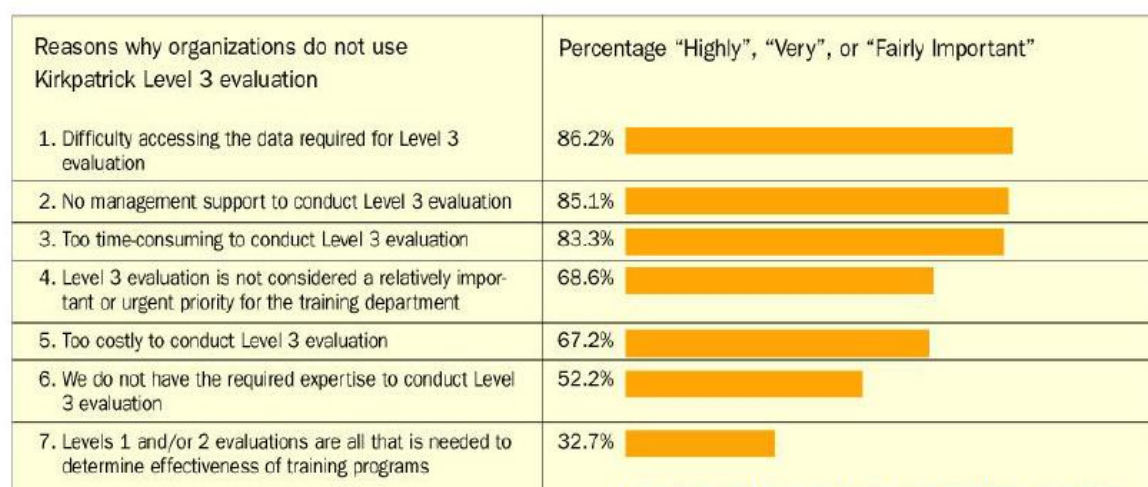
Reasons why organizations do not use Kirkpatrick Level 3 evaluation	Percentage "Highly", "Very", or "Fairly Important"	
1. Difficulty accessing the data required for Level 3 evaluation	86.2%	
2. No management support to conduct Level 3 evaluation	85.1%	
3. Too time-consuming to conduct Level 3 evaluation	83.3%	
4. Level 3 evaluation is not considered a relatively important or urgent priority for the training department	68.6%	
5. Too costly to conduct Level 3 evaluation	67.2%	
6. We do not have the required expertise to conduct Level 3 evaluation	52.2%	
7. Levels 1 and/or 2 evaluations are all that is needed to determine effectiveness of training programs	32.7%	

Figure 2. Survey results on barriers to trainee behaviour and results. Adapted from ''Usage and Value of Kirkpatrick's Four Levels of Training Evaluation'' by J. Pulichino, 2007, Doctoral dissertation, p. 90.

2.4.5 Time and Frequency

The final basic criteria within an evaluation is to determine the moment of evaluation and the frequency of the evaluation. There is little research on the frequency and time of evaluation. However, Wick et al. (2010) suggest that there should be a period of time between the training and the evaluation for the newly acquired knowledge, skills and attitudes to endure. A behaviour change needs time to be visible and the organisation should give the learner time and opportunities to show the posttraining behaviour in practice (Arthur, Bennet Jr., Edens, & Bell, 2003).

All the mentioned criteria are seen within this research as important elements in designing an evaluation approach and they may vary in a particular organisation or situation and therefore are dependent regarding the context. For this research, the focus is on evaluating on behaviour level, and therefore there is chosen to identify relevant characteristics for the design of an evaluation approach on behaviour level.

3 Research question

This study investigated the area of training effectiveness evaluation. The aim of this study was to investigate what evaluation characteristics are seen as relevant by learning professionals in designing an evaluation approach on behaviour level with the goal to determine training effectiveness. By going back to the literature and existing evaluation frameworks that are used to guide an evaluation, there is

explored what these characteristics mean for the current literature and the existing evaluation frameworks.

Summarized, the research question that is posed within this study is: '' What evaluation characteristics are seen as relevant by learning professionals in designing an evaluation approach on behaviour level to measure the effectiveness of training programs and what does this mean for the existing evaluation frameworks?''

4 Method

4.1 Research design

This research investigated what evaluation characteristics are seen as relevant in designing an evaluation approach and the implications for the current literature and evaluation frameworks. A case study, generating qualitative data, is used to investigate this topic. This design is chosen, because a) the study answers a 'what' question, b) the behaviour of the respondents cannot be manipulated, and c) the context of this study is relevant for this research (Yin, 2003). The situation is chosen for Deloitte specific and therefore a case study is an appropriate design. It can be seen as a preliminary study that strives to give insight how organisations can evaluate trainee behaviour and results. This research has a pragmatic paradigm. It is oriented 'toward solving practical problems in the "real world" (Feilzer, 2010, p8) and the collection of qualitative data provides a more in-depth view and perception of the employees of learning professionals.

4.2 Organisational context

This study is conducted in the organisational context of Deloitte. Deloitte is an organisation that employs professionals in independent firms that are globally located to provide Audit, Tax, Consulting, Financial Advisory and Risk management to clients. In order to keep their employees competitive, Deloitte provides training opportunities for all employees in every function. The learning department of Deloitte exist of several areas. The design and development team is responsible for the design and development of the internal training programs and e-learning. A team of learning advisors, one for every function, is responsible for the learning needs analysis and the translation of the educational wishes from the functions into a curriculum. Another important group within the learning department is the group of program managers, who are in charge of the larger programs and are the link between the learning advisor, the learning contact from the function and the event management. Finally, the event management team is responsible for the planning of the curriculum. Their main activity is to schedule the curriculum and make sure that the administration regarding the training programs is up to date.

The employees of Deloitte are the 'product' of Deloitte, meaning that their knowledge and skills directly contribute to the business outcomes of Deloitte. Therefore Deloitte invests a lot in training programs. Deloitte offers learning opportunities according to their vision: they use the 70:20:10 method, which means that 70% of the learning happens on the job, 20% in contact with peers and supervisors and 10% in a formal training context. This 10% training is organised in a formal classroom setting. In order to make decisions on the continuation or adaption of the offered programs Deloitte wishes to know how effective the provided training programs are. To measure this, the learning department of Deloitte currently evaluates right after the training by measuring the reaction of the trainee by an online evaluation right after the training. This evaluation focuses especially on the design and content of the training and the role of the facilitator. Within the evaluation there is also asked if the trainee thinks that the learning knowledge and skills will be applied on the job. However, if the learning department wishes to determine whether the offered internal training programs were effective, the learning department needs a structured approach to evaluate. This is currently absent. Decisions about the continuation or adaption of a training program are within Deloitte currently done based on the amount of cancellations, the evaluation of the trainee right after the training, the opinion of the facilitator about the training and budgetary information

4.3   Participants

The participants of this research are learning professionals of the internal learning department of Deloitte Netherlands. The department counts 20 employees. The sampling technique that is used for this research is ''stakeholder sampling'' (Palys, 2008). This type of purposive sampling is most useful when a policy should be analysed and it concerns the identification of the important stakeholders who are related to or affected by the program that is researched. A stakeholder analysis identified the most important stakeholders and decision makers of the learning department. Inclusion criteria for this research were:

1) Participant had to be a member of the learning department.
2) Participant had to be a learning advisor, program manager, event manager or part of the design and development team.
3) Participant should have a role in the decision-making process or evaluation process.

By including these respondents based on these criteria, the whole learning department is included in this research. Exclusion criteria for this research is that employees of the learning department that are not a learning advisor, program manager, event manager of design and development are not invited to participate in this research. Because of the scope of this research there is chosen to exclude respondents that are not related to the learning department of Deloitte and who do not have a role in the evaluation process. In total 18 members of the learning department were approached to participate in the focus-group interviews. 14 participants actually participated in the focus-group interviews. These participants were divided in three groups, based on their function. In every focusgroup there was included a learning advisor, a program manager, an event manager and design and developer.

For the overall group, 14, 3% of the participants is male and 85,7% is female within the age range varying from 26 to 50 years old (M=36.14, SD= 6.5). 35,7% of the participants live in Rotterdam and satellite towns, 35,7% live in Amsterdam or surrounding satellite towns, 14,3% in The Hague or satellite towns and the remaining 14,2% lives in Utrecht or other places in the Netherlands. The educational level has been subdivided into Secondary Vocational Education (7,14%), Higher Vocational Education (57,14%) and University Level (35,7%). 28,6% of the participants work 3 to 5 years at Deloitte. The work experience within learning and development field of the sample varies between one to three years and 10 to 15 years.

Within the first focus-group interview, 4 female and 1 male respondent participated. The age range within this group is between 28 and 39 years (M=34.4, SD= 4.67). 3 program managers attended, 1 event manager, 1 design and developer.

For the second focus-group interview, there attended 4 participants and they were all female. The age varied between 26 years and 39 years (M= 34.00, SD=5.60). Within this interview 1 learning advisor attended, 1 program manager and 1design and developers and one manager of the event management team.

Within the final focus-group interview 4 female participants attended and 1 male participant. The age varied between 30 years and 50 years (M=39.6, SD=8.29). 2 learning advisors attended, 1 program manager, 1 design and developer and 1 event manager.

4.4   Procedure and instrumentation

The focus-group interviews had a duration of one hour and were conducted during December 2015 and January 2016. In total, three focus-group interviews were held. Participants were invited for the focus-group interviews via e-mail. The participants were selected on the previously mentioned inclusion criteria. Before participating in the focus-group interviews, participants were asked to give written consent. The qualitative focus-group interviews were used to gather data within this research, because it is beneficial to gather participants' perceptions and feelings about certain topics and the interaction between participants may generate important data in a short period (Onwuegbuzie, Dickinson, Leech,

& Zoran, 2009). This context required a focus-group interview for the explicit use of group interaction and the development of ideas on defining an evaluation approach.

The focus-group interview were semi-structured. In order to keep the interview semi-structured, a topic list (figure 3) was created with the most important constructs (Bates, 2004). The input for the topic list in the interview is retrieved from the key questions that should be asked when designing a program evaluation according to Zinovieff (2008) and the practical paradigm of Hashim (2007).
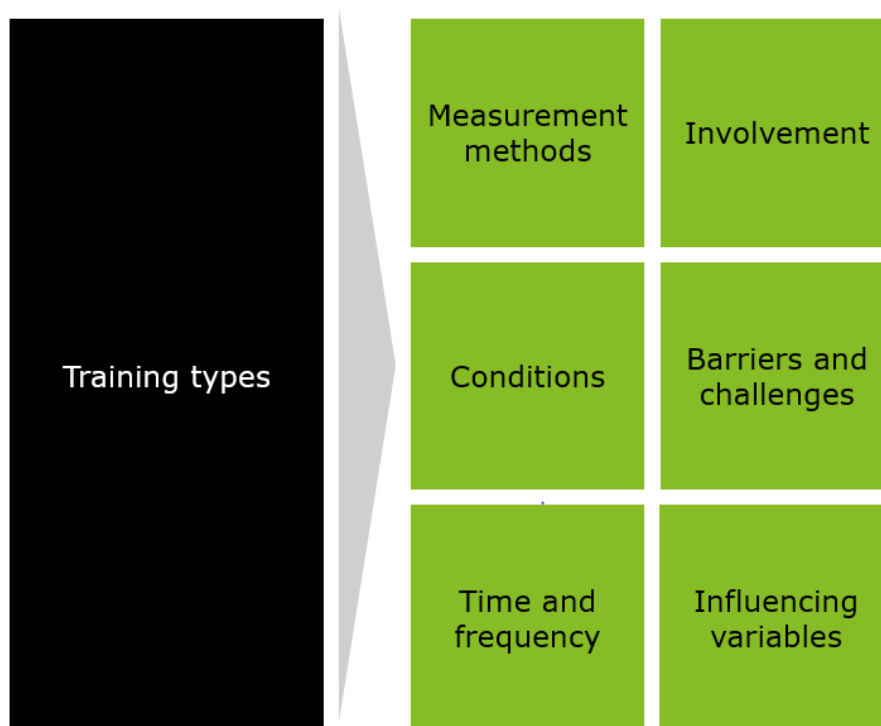

Figure 3. Topic list for focus-group interviews

First, respondents are asked to identify the training types that are offered within Deloitte. This question is asked, because in order to design an evaluation approach, the type of training is the base for the measurement methods that are chosen. This is followed by the 6 pillars and basic evaluation criteria.

The topic list gave guide to the discussion and brainstorm. There is chosen to discuss topic instead of asking fixed questions, to avoid steering from the moderator and gather as much data as possible. The questions that were asked in the focus-group interviews were open-ended. According to literature, a successful focus-group interview contains 6 to 12 participants. However, Onwuegbuzie et al. (2009) state that 3 or 4 'mini focus-groups' can be helpful when participants have a specific knowledge or experience. Because in this study there were different teams who collaborate with each other and have specific knowledge, an amount of 3 to 4 participants is chosen and in total 3 focus-group interviews.

The moderator, in this study the researcher, introduced the topic and assisted participants in a discussion, and encouraged interaction. The researcher guided the conversation with help of the topic list. The researcher listened and avoided personal opinions to avoid influencing the participants. The researcher and the participants were in direct contact during the research period and there existed a relationship between the researcher and the participants, because the researcher is located in their workplace for six months. However, both parties have separate activities and are not dependent of each other. After the focus-group the participants are asked via e-mail to anonymously fill out a short survey on demographic variables (appendix B).

4.5   Data analysis

The focus-group interviews were recorded and transcribed manually. Transcribing the interviews ensured that the context and interpretation of the participants could be taken into account (Bates, 2004). A constant comparison analysis was used to analyse the data resulting from the focus-group interviews. This analysis contained three stages (Onwuegbuzie, Dickinson, Leech, & Zoran, 2009). Within the first analysing phase data was separated into small units, which is referred to as open coding. Within this stage the units in the column 'code' and 'code description' were identified, based on the literature study.

Subsequently, axial coding was used and the codes were assigned to a category. Axial coding is applied, because it allows theory-driven coding and data-driven coding as well (Boeije, 2008). The main categories that were defined can be seen in Appendice A. Almost all categories were theory driven, except for the categories conditions and time and frequency. These were data-driven codes. Figure 4 is an illustration of the axial coding process that is used in this research.. Emerging themes within the three different focus-group interviews were identified by an across-group analysis (Onwuegbuzie et al., 2009).
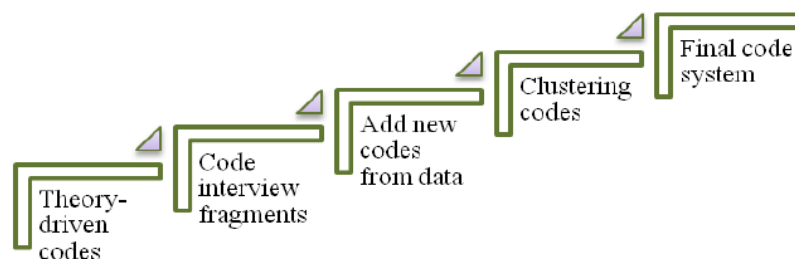
Figure 4. Axial coding process. Adapted from *Analyseren in Kwalitatief Onderzoek: Denken en Doen. (p.98)*, by Boeije, 2008, Amsterdam: Boom Lemma

Construct validity was increased by letting a key informant read the case study report, as mentioned by Yin (2003) as a way to increase the construct validity and reliability. Cohen's κ was run to determine if there was agreement between two raters on whether 103 fragments were coded right. After several conversations, there was substantial agreement between the two raters judgements, κ = .805. There should be taken into account that there could be a response bias or a form of reflexivity, the participants might have answered what the researcher wants to hear. Internal validity is ensured by pointing out patterns, similarities and differences between participants' perceptions and the theoretical framework. External validity is assured by describing the coherence with the theoretical framework. A presentation for the learning department was organized to inform the learning department about the research results, conclusions and recommendations.

5   Results

Fist, this study explored what evaluation characteristics are seen as important in designing an evaluation approach on behaviour level and what this implicates for the theory and current literature. The professionals of the learning department of Deloitte provided an answer on this research question based on the focus-group interviews. Table 2 shows a quantitative overview of statements that have been made within the focus-groups interviews.

Table 2
Quantitative overview of statements

Topic	Focus-group 1	Focus-group 2	Focus-group 3	Total
Training types	38	6	30	74
Involvement	25	30	30	102
Measurement methods	78	53	47	195
Barriers	38	43	64	125
Conditions	29	81	44	169
Time and frequency	10	12	19	41
Other/ not relevant	101	53	69	223
Total	319	278	332	929

The data that is retrieved from the focus-group interviews did not allow a quantitative review of the results because of the small sample, therefore a qualitative explanation and interpretation of the results has been given.

Training types
Within this category the respondents identified what types of training Deloitte offers and what these types of training say about the possibilities and focus than organisations can have in evaluating. The respondents make a distinction between several categories. A distinction can be made between trainings based on content, trainings based on size or effort, and trainings that are provided without the involvement of the learning department.

The first topic that is mentioned within the all the interviews are the trainings that are based on content. Within all three interviews, respondents agreed that the offered internal trainings programs can be distinguished, based on content, in two categories: skill trainings and knowledge trainings. One respondent stated that that soft skill trainings are more important in the current business of Deloitte. One focus-group concluded that skill training can be evaluated on behaviour level, whereas knowledge training is more difficult to evaluate on behaviour level, because it is difficult to observe if a change in behaviour is due to the newly acquired knowledge.

> *I: '' So I understand: when you want to evaluate on knowledge training to see if they are effective and a behaviour change has occurred, that it is difficult to do that?''*
>
> *R: '' Yes''.*
>
> *R:'' It is mostly about skills''* (Appendix C, 3.29, Manager and interviewer)

A second category that has been distinguished is that the training programs can be evaluated based on the size and effort. There was no consensus between the respondents in this category. Within the second focus-group interview the question is raised by 2 respondents that they do not know if the amount of effort that is put into evaluating is worth the results. Therefore they mentioned that small trainings, such as a PowerPoint training is a low priority. More expensive programs are more interesting for this group to evaluate. This in contrary to the respondents of the first interview, who were interested in the soft skill training evaluations. A suggestion that is done in one focus-group interview, is to

categorize the training programs into programs with a low relevance to evaluate, a higher relevance and a selection of programs that are urgent to evaluate on.

A final category is that Deloitte offers training programs that are not initiated by the learning department and they are also not involved in the creation of some training programs. These trainings are mostly related to the Consulting practice and are organised by the business. Therefore these training programs are not seen as urgent to evaluate by two focus-group interviews. This topic has not been mentioned in the first focus-group interview. A second element that is mentioned in two focus-group interviews is that there are training programs that are offered by the learning department, but the content and quality of the training is in hands of an external institute, in Deloitte's case this is the Nederlandse Orde van Belastingadviseurs (NOB) or the Autoriteit Financiele Markten (AFM). For this trainings the respondents stated that they do not feel the urge to evaluate these trainings.

In conclusion, within Deloitte there are three training types that can be distinguished: soft skill trainings, knowledge trainings and trainings in which the content is fixed by an external party,

Measurement methods

Within this category several methods to evaluate and measure a behaviour change are distinguished. The measurement method that is mentioned in all three interviews is the use of a performance manager. A performance manager is a manager that coaches a counselee and observes the performance of the counselee. Performance managers seem suitable for monitoring and observing the counselees on behaviour aspects. In two of the three interviews is mentioned that the learning department could interview the performance manager in a formal telephone call and ask if he or she observes a behaviour change as a result of the training.

'' If we want to measure that, we should ask the performance managers now and then to ask: '' Hey, do you see results of the training'' (Appendix A, 1.224, Design and Developer)

The respondents of the first focus-groups mentioned that at this moment there are no behaviour indicators created for the existing trainings. For this group it seems like a solution that a training design should include an evaluation plan with characteristics and that a performance manager can observe and determine whether a trainee has had a behaviour change.

A second measurement method that is mentioned in two out of three interviews, is sending out a survey to the trainees and ask their own self-report on their changing behaviour. One respondent mentioned a negative argument to this way of self-reporting is that it is subjective and that an e-mail can 'quickly be ignored and deleted' (Appendix B, 2.168, Design and Developer). It is less personal than an interview. However, one respondent stressed that this is a less time-consuming way for the learning department and the business

A third measurement method proposed in all three focus-group interviews is to integrate the aspect of behaviour change in the performance assessment. Every employee of Deloitte is assessed based on performance in a mid-year interview and an end-year interview, in which behaviour plays a central role and is worth 50% of the performance interviews. In every interview there is suggested to link the performance assessment with the trainee's targets and to ask feedback from peers or clients.

> *'' Make it a part of the perform cycle. I want to follow this training because…. I have this target and I want to work on this behaviour and with the mid-year interview you can refer to it… It doesn't have to be done via iValue but receive feedback specifically on behaviour and ask colleagues, but that can also be a client'' (Appendix B, 2.83, program manager).*

By including peer and client feedback, there are multiple sources that can observe a behaviour change. 2 out of 3 interviews mentioned a 360 degree feedback as a possible way. A negative argument that was mentioned by 1 respondents regarding this method, is that not every training allows the formulation of concrete and measurable behaviour indicators.

Other measurement methods that were mentioned are the use of a control group or a peer group comparison. In one of the focus group interviews is mentioned that one group can follow a training and

another group functions as a control group. A performance manager monitors the groups and provides this information. An argument that is proposed for this measurement method is '' that it is very time-consuming'' (Appendix A, 1.397, Design and Developer).

In short, the three measurement methods that are proposed in the interviews are the monitoring and observing by performance managers, self-reports by the trainees in the form of a survey, the integration of the evaluation in the performance assessment by using 360 degree feedback, and control-groups.

Involvement

The respondents reported positive and negative arguments regarding their involvement in the measurement and evaluation process. In two out of three focus-group interviews the argument is used that they question who should be responsible for the measurement of trainee behaviour. There was no consensus between the respondents. In the first interview, 1 respondents mentioned that that the measurement of trainee behaviour should not be the task over the learning department. The other respondents from the first interview contradicted that it is the job of the learning department to provide a good training for a problem and therefore feel responsible to see if the training was a solution for the problem. Next to that, a remark is made by one of the focus-group interviews is that if the learning department wishes to know if their offered programs were effective, they should also take initiative in this. Occasionally they receive feedback on the effectiveness of programs in an informal way, but do not make a documentation of this and use it for decision-making. The most important person who should be involved in the collection of data and the measurement of behaviour according to the respondents in all three interviews, is the performance manager. The performance manager is seen as the most suitable person to identify a behaviour change for a group of respondents.

> '' *Then the performance counsellor can be used. It would be nice if in all the mid-year interview and end-year interviews it would be a mandatory topic for the performance manager to discuss: I see in the system that you did this training, make clear what this did for you. How did your behaviour change? Is it a desired behaviour change or is it a behaviour change that we had not foreseen?''* (Appendix B, 2.16, manager)

Summarized, the learning department has different views on their role in the evaluation process. The majority of the learning professionals thinks it is the job of a learning department to take initiative in collecting data. However, they all agree on the performance manager as the most suitable person to collect data from.

Conditions and organisational characteristics

Within this sections there is discussed what organisational characteristics Deloitte has and how they have influence on the evaluation approach. The respondents reported several important conditions that need to be met before an evaluation on behaviour level can be executed. The first condition that is mentioned in all three focus-group interviews is that if the learning department wishes to evaluate, the collaboration and involvement of the business should be enhanced. According to two of the three focus-group interviews, this requires a switch in mind set and commitment of the whole organisation.

> '' *If they don't find it important, than there is less commitment I think''* (Appendix B, 2.145, Design and Developer).

The second condition that is mentioned in one of the three focus-group interviews, is that if behaviour plays a central role, there could be formulated quantitative and measurable indicators for behaviour earlier in the design process. Only then it is possible for the performance manager to assess

the behaviour change in a structured way. This also means that the learning department and the business should speak the same language and that a lot of knowledge is asked from the performance manager.

'' And what you want to really evaluate must be determined at the beginning as one of the targets. You can hardly evaluate what not… you have to make it very concrete in the objectives you want to evaluate. And for example in Negotiation skills,, one of the objectives is that people know that a model exists. You could say, after this training, for example: someone must have used at least three parts of that model, and that is something that can be checked. So the more concrete you are in the beginning, the easier it is to evaluate at the end. '' (Appendix C, 3.70, manager).

Next to that, two respondents mentioned that when a behaviour change is investigated, it might be that there are other factors that have influence on the behaviour change and that a behaviour change might be the result of something else, but not a result of training (e.g. a good mentor, an article that has been read). On the other hand, respondents stated that trainees do not always get the opportunity to show a different behaviour.

One respondent stated that not only the behaviour indicators should be determined, but the goals and the objectives should be clear as well. Therefore a communication with the business is described as helpful. A last important condition that is emphasized in every focus-group interview is that time and budget are elements that should be available to evaluate on behaviour level.

''The conditions should be created that somebody gets the opportunity to do so.'' (Appendix C, 3.158, Design and Developer).

In conclusion, the respondents state that first a collaboration and involvement with the business is important. Second, there should be formulized measurable behaviour indicators for performance managers to observe, and time and budget should be available.

Time and frequency
Regarding the time and frequency of the evaluation, all three focus-group interviews reported that they do not think that evaluation of these programs should be done often. One focus-group thinks that once a year is enough, the other two focus-groups see twice a year as enough to have insight in the effectiveness of their offered programs. One of the respondents of the second focus-group interview mentioned that according to neurosciences behaviour changes after 40 days into a habit. Other respondents mentioned that 3 months after the training would be a suitable moment to collect information from the performance manager.

*''If he or she still shows that behaviour after three of or four months, it is assumed that it will be shown afterwards as well. '' (*Appendix C, 3.336, manager).

In conclusion, respondents stated that once a year is enough to evaluate, preferably 3 months after the training.

Barriers and challenges
Within this section, respondents listed several barriers for the evaluation on behaviour level. The main barriers that are discussed in all three focus-group interviews are about difficulty in accessing data, a lack of time and the relevance of using trainee behaviour as a predictor instead of trainee reaction. Within the third focus-group interview respondents stated that it is difficult to access data on behaviour level, especially without the management support.

'' Of course there are business outcomes and learning goals et cetera, and you can evaluate that, but skills are not always seen in practice that clear. That someone sits next to you and checks: yes this is

going well. It is also not always visible the next day in practice because it is much more for the long term'' (Appendix C, 3.63, program manager)

Besides that, in two of the three focus-group interviews the respondents agreed on that it the learning department does not have the right knowledge or tooling to gather this data and therefore the employees in the business should be involved.

Another perceived barrier in all three focus-group interviews is time. In every interview there is mentioned what the benefits of evaluation on trainee behaviour are, but this means spending time for the learning department to invest in the training evaluation, but also asking time from employees. This can be in the form of a self-report or an observation, but both methods cost a lot of time. Within two of the three focus-group interviews respondents disagreed about the relevance for the learning department when it comes to behaviour level. For two respondents the reaction evaluation seemed sufficient. One of the respondents claimed that information of changing behaviour is not relevant for the learning department, but more beneficial to the performance manager of a trainee. Two respondents claimed that when there is lack of a behaviour change, the learning department should reconsider if the program is suitable and the training might be adapted.

Summarized, respondents mentioned the following barriers: difficulty in accessing data, the lack of time and the relevance of using trainee behaviour as a predictor instead of trainee reaction.

6   Conclusion and discussion

The purpose of this study was to investigate what evaluation characteristics are seen as relevant by learning professionals in designing an evaluation approach and the implications for the theory on training effectiveness. There is taken into account what types of training allow an evaluation on behaviour level and what are possible measurement methods. Next to that, there is explored who should be involved in this process and what conditions should be met in order for organisations to evaluate. Finally, there is investigated what possible barriers and challenges are for the evaluation of training effectiveness on a behaviour level. Below, results of this study are summarized and discussed. The first purpose of this study was to investigate what evaluation characteristics are seen as relevant by learning professional in designing an evaluation approach in the context of an accountancy multinational. Results showed that there are several characteristics and elements seen as important in designing an evaluation approach by learning professionals in practice.

The first finding is related to the type of training that Deloitte distinguishes. Respondents in every interview reported that the formal internal trainings programs can be distinguished into two main categories: skill trainings and knowledge trainings. The respondents claimed that the evaluation of behaviour regarding knowledge training is difficult, because knowledge is not easily translated into behaviour. This is in line with the research of Kraiger et al. (1993), who stated that knowledge trainings can be best evaluated by recall testing and multiple-choice or true-false questions. Although these measurement methods are not specifically mentioned by the respondents, there is stated that they perceive difficulty in evaluating knowledge training on behaviour level. When it comes to the skill trainings, respondents mentioned that skill trainings are more suitable to be evaluated on behaviour level. Kraiger et al. (1993) found the same finding, because they states that '' in many domains, it would be useful to track trainees' skill development to assess progress and to design or modify other training interventions'' (p.317). These training types are not an evaluation characteristic, but every type of training needs a different evaluation approach.

Besides the distinction that organisations make between the types of trainings that they offer, outcomes also indicated that organisations do not see the urge in evaluating on every program, but that they might make a selection of programs. This selection could be based on the division of large or expensive programs, but there are also programs within Deloitte's curriculum that cost a lot of effort. Results showed that organisations might prefer to evaluate complete programs instead of regularly offered separate trainings. There are also trainings that are offered by the learning department, but the

content and quality of the training is in hands of an external institute. Literature does not say anything about this topic, however it is assumed that the learning department of Deloitte should be responsible to decide whether this has a high priority or not. Every company has its own reasons to evaluate and literature does not mention what program should be a top priority for organisations. When looking at the IMTEE, the model does not take into account what types of training programs can be evaluated when looking at the measurement level of transfer performance. This means that the model provides a general guideline, but does not specify the possibilities for specific training programs.

The IMTEE or other evaluation models and frameworks do not include proposal or guideline for measurement methods. In the process of designing an evaluation approach, organisations wish to know what types of measurement methods are suitable within the context of their organisation. For Deloitte, respondents reported several methods to evaluate and measure a behaviour change. The measurement method that is mentioned most often and seemed most suitable according to the respondents, is the use of performance managers who monitor and observe the trainees on behaviour aspects that are described in a learning needs analysis. AlYahya and Norsiah (2013) stated in their research that an observation is a suitable tool measure behavioural changes. Kraiger et al. (1993) stress as well that the measurement of changing behaviour and in particular skills, are possible when targeted behavioural observations are used. Next to that, when learners know that they are being monitored and receive feedback, the transfer performance after training will likely increase (Velada et al, 2007l; Wick et al., 2010). Therefore observation and monitoring of a performance manager can be seen an evaluation characteristic which might influence transfer performance.

A second measurement method that is suggested by the respondents, is sending out a survey to the trainees and ask their own self-report on their changing behaviour. Respondents have several negative arguments, for this, because it is a self-report and is less personal than an interview. However, they think that it is less time-consuming for the learning department and the business does not have to be involved. Sachdeva (2014) mentioned in her literature review that a written evaluation allows for more meaningful responses and might contribute to the determination of the effectiveness of the training. Sachdeva (2014) adds to that, that by sharing their experiences it might be that some issues are addressed which otherwise might have been unnoticed. This means that this measurement methods seems more suitable than expected of the employees. However, with a survey the chances of bias is larger, because trainee are less objective. Donaldson and Grant-Vallone (2002) stated in their research that research participants respond in socially desirable ways and are likely to under-report or over-report their behaviour, depending on the situation. This seems a problem according to Donaldson and Grant-Vallone (2002) in particularly in organisational behaviour research, because employees are sometimes afraid that their responses might jeopardize their job. Therefore Deloitte could take into account that this measurement method is not reliable on its own. Borman (1991) concluded that peer and supervisor ratings were more accurate than self-reporting. This statement is done in the context of performance rating, but it is assumed that this form of bias is also applicable in this research.

A third suggested measurement method is the integration of evaluation with a performance assessment. Respondents proposed to link the performance assessment with the trainee's targets and to ask feedback from peers or clients and include a 360 degree feedback. This is in line with research of AlYahya and Norsiah (2013), who stated that a supervisor that observes behaviour might give bias. By including peer and trainee feedback, the bias will be reduced and a better example will be given of the behaviour change, as is also concluded by Borman (1991). However, Jellema, Visscher and Scheerens (2006) contradict this and concluded that a 360 degree feedback is not seen as an efficient training evaluation method. This means that the balance must be struck whether the use of a performance manager only is sufficient for the learning department to base their conclusions on. Research does not confirm the argument that not every type of training allows the formulation of concrete and measurable behaviour indicators. Additionally, Velada et al. (2010) found that when trainees know they will receive feedback on their performance after training, transfer performance will increase.

Other measurement methods that were proposed are the use of a control group or a peer group comparison. It is suggested that one group can follow a training and another group functions as a control

group. However, Habicht et al. (1999) state that decision-makers use a control-group as a method to rule out external factors. Doing so, will result in making more plausible conclusions. For the learning department of Deloitte this might seem a bit too elaborate to determine the effectiveness of the training programs. This is also, because conclusions can be drawn wen the control group is similar to the intervention group. This is a requirement that is also proposed by Kirkpatrick (1994). Therefore, it seems a possible measure method when there is a program that emphasizes on ruling out external factors and when this is practically possible. However, for a large company with different units, as is the case with Deloitte, this measurement method is not recommended.

Regarding the involvement within the evaluation process, the IMTEE or other frameworks do not propose who should be involved in the evaluation and measurement process. Results from this study showed that the learning department, performance managers, peers and clients could be involved in the measurement process. Respondents stated that if the learning department wishes to know if their offered programs were effective, they should also take initiative in this. Literature does not specifically mention who should take the initiative. It is however assumed that if the learning department is a decision-maker regarding the training programs, that they will be the initiator to gather data regarding the effectiveness of training. The most important person who should be involved in the measurement of behaviour according to the respondents, is the performance manager. The performance manager is seen as the most suitable person to identify a behaviour change for a group of trainees. This is in line with Sachdeva (2014), who stated that if desired behaviour is sought, an immediate supervisor could observe this behaviour. AlYahya and Norsiah (2013) contradict this statement, because according to them a supervisor only could be biased. Borman (1991) adds to this that the involvement of peers and self-reports as a combination decreases the bias.

When a survey is used as a measurement method and sent to the trainee, it seems logical that the trainee should be involved. However, as mentioned in the previous section, a self-report might cause socially desirable answers. To avoid this, a combination of involvement of self-report, supervisor and peers might give a good indication of the behaviour change and the effectiveness of training. Kirkpatrick (1994) also creates some guidelines on evaluation and also suggested that an evaluation should have multiple sources, such as the trainee self, his or her manager, a subordinate and others who may observe the participants' behaviour. Jellema, Visscher and Scheerens (2006), as stated in the previous section, do not confirm this statement and concluded that co-workers do not perceive a behaviour change.

One main concern for the learning department is that the chance exists that other factors have influence on the behaviour change and that the results cannot be linked to the training but might have a different cause. As mentioned before and stressed by Pulichino (2007) and Habicht et al. (2004) , a control group can be used to rule out any external factors, but has the disadvantage that the experimental group and the control group should be exactly equal, which is not the case in most situations. Therefore Deloitte might inventory what has their preference: the creation of a control group which costs a lot of time, or making conclusions based on the findings that might be not fully reliable.

The IMTEE includes organisational characteristics as an influencing variable. The most influencing characteristic, according to Alvarez et al. (2004) is the positive transfer environment, which is defined by Alvarez et al. (2004) as ''an environment that encourages trainees to implement skills learned in training such as supervisory support, rewards for applying learned skills, and follow-up performance evaluations'' (p.403). An organisational characteristic about the environment that is mentioned, is that the organisational culture of Deloitte sees learning as a 'magical solution' and the influence of the learning department is not very high. A good understanding between the business and the training department could be beneficial to tackle this challenge (Pulichino, 2007). On the other hand, respondents stated that trainees do not always get the opportunity to show a different behaviour. Kirkpatrick (1994) has two solutions for this: convincing the managers that evaluation has a key role in their business and proof that evaluation on behaviour level will results into a better transfer. A second solution is to inspire leaders to participate and be active in the training, to convince the importance of

training. The respondents of this interview confirmed that a positive transfer environment is essential in the evaluation process.

Respondents reported other conditions within in an organisation that need to be met for the learning department to evaluate on behaviour level. Firstly, if a learning department wishes to evaluate, the collaboration and involvement of the business should be enhanced. According to them, this might require a switch in mind set and commitment of the whole organisation. Pulichino (2007) pointed out within his research that collaboration between the learning department and business units is necessary, as well as a challenge. This is confirmed by the respondents for the case of Deloitte.

The second condition that is mentioned by the respondents, is that if behaviour plays a central role, there could be formulated quantitative and measurable indicators for behaviour earlier in the design process. Only then it is possible for the performance manager to assess the behaviour change in a structured way. This also means that the learning department and the business should speak the same language and that a lot of knowledge is asked from the performance manager. Again, collaboration is the key word in this case, when taking into account what Pulichino investigated. The business unit might have a better view on the required behaviour and might be more competent in establishing the behaviour indicators. Pulichino adds to this in his research that bringing the training departments into the workplace is an option. However, for this case it seems more logical to let the business establish the desired behaviour and translate them into observable behaviour indicators.

A last important condition that is emphasized by the respondents, is time and budget. Pulichino (2007) suggests that organisations should delicately plan how they will collect the data and take into account what time constraints are applicable to the situation. This seems an appropriate approach for organisations. However, it might be beneficial that the role division with the business units and a learning department are clearly described before making a planning. Collaboration and involvement of the business, having the same vision as the business and time and budget are factors that are not included in the IMTEE. However, these elements are seen as important characteristics by the learning professionals.

Regarding the time and frequency of the evaluation, respondents reported that they do not think that evaluation of these programs should be done often. Once or twice a year is seen as enough to have insight in the effectiveness of their offered programs and 3 months after the training would be a suitable moment to collect information from the performance manager. Kirkpatrick (1994) suggests with his guidelines that if it is possible, an evaluation before and after training is beneficial to observe a change. However, when taken into account that the learning department does not have the time and resources for this type of evaluation, this does not seem an appropriate method for Deloitte. Evaluating after the training is more convenient. However, Kirkpatrick (1994) does mention that before measuring the behaviour change, an essential aspect is to allow some time for the application of the new knowledge and skills on the job. The new behaviour will not be visible right away. This also a finding within this research. Therefore the determination of an appropriate time should be found. Kirkpatrick (1994) emphasizes that the exact moment depends on the organization and the situation. Mostly, a first evaluation can be done two or three months after the training. A second moment can be done 6 months after that. As the respondents also suggested an evaluation twice a year, this seems a logical choice. Within the IMTEE, there are no suggestions done regarding the moment and frequency to evaluate according to the framework. This is important for an organisations, to determine what the best moment to evaluate is.

Respondents reported a number of barriers for the evaluation on behaviour level. The main barriers that are reported by the respondents are about difficulty in accessing data, time and the relevance of using trainee behaviour as a predictor instead of trainee reaction. The respondents state that it is difficult to access data on behaviour level, which is also the primary reason for organisations not to evaluate on behaviour level according to the research of Pulichino (2007). Management support is mentioned by Pulichino (2007) as an important factor, which is in line with the findings of this study.

Another challenge that is reported by the respondents, is that the learning department does not have the right knowledge or tooling to gather this data and therefore the employees in the business should be involved. There is an explanation for this, as stated by Pulichino (2007): evaluations on

reaction level are more related to the training professionals, in contrast to measuring behaviour that happens on the job. A requirement mentioned by Pulichino is that a collaboration across the organization is beneficial for evaluation on behaviour level.

Time is a perceived barrier for the respondents as well. The respondents see the benefits of evaluating on trainee behaviour, but this means spending time for the learning department to invest in the training evaluation, but also asking time from employees. This can be in the form of a self-report or an observation, but both methods cost a lot of time. Pulichino has an explanation for this lack of time: behaviour is complex to measure and therefore costs time to find out. Next to that, time is seen as the most mentioned reason and defeats cost or lack of expertise. Kirkpatrick (1994) reported a suggestion on how to challenge this barrier: by choosing a few typical trainees and interview or survey them. An organisation may assume that if they have changed their behaviour, the rest of the trainees have done that as well. When organisations do not have enough time to invest in evaluation, this might be a solution.

Respondents also reported mixed findings on the relevance for the learning department when it comes to behaviour level. For some trainings the reaction evaluation seemed sufficient. Other respondents claimed that when there is lack of a behaviour change, the learning department should reconsider if the program is suitable and the training might be adapted. Pulichino found that this barrier is not seen as the most important one, but is a sign of priority division between the learning department and the business units they help. It is mentioned that it might be that learning professionals are contented with the measurement that they have in control (such as the evaluation on reaction level, in the form of a survey). Pulichino suggests that training departments might benefit from an experience in the real workplace to see how behaviour is embedded in the day-to-day job in order to conduct a good evaluation.

The barriers are not taken into account in the IMTEE, but seem important to include in the evaluation framework. These barriers are experienced in practice and play a role choosing and designing an evaluation approach. Some measurement methods or the involvement of people depend on the resources and the possibilities. Therefore these barriers might influence the transfer performance.

In conclusion, the main evaluation criteria (measurement methods, involvement, conditions, time and frequency, and barriers and challenges) and the identified evaluation characteristics when to evaluate on behaviour level (figure 4) are seen as relevant in designing an evaluation approach on behaviour level by learning professionals in an organisation for professional services. These characteristics, which are critical in designing an evaluation approach, are currently not taken into account in evaluation and effectiveness frameworks. In the next paragraph there is explained what this means for the existing evaluation frameworks.
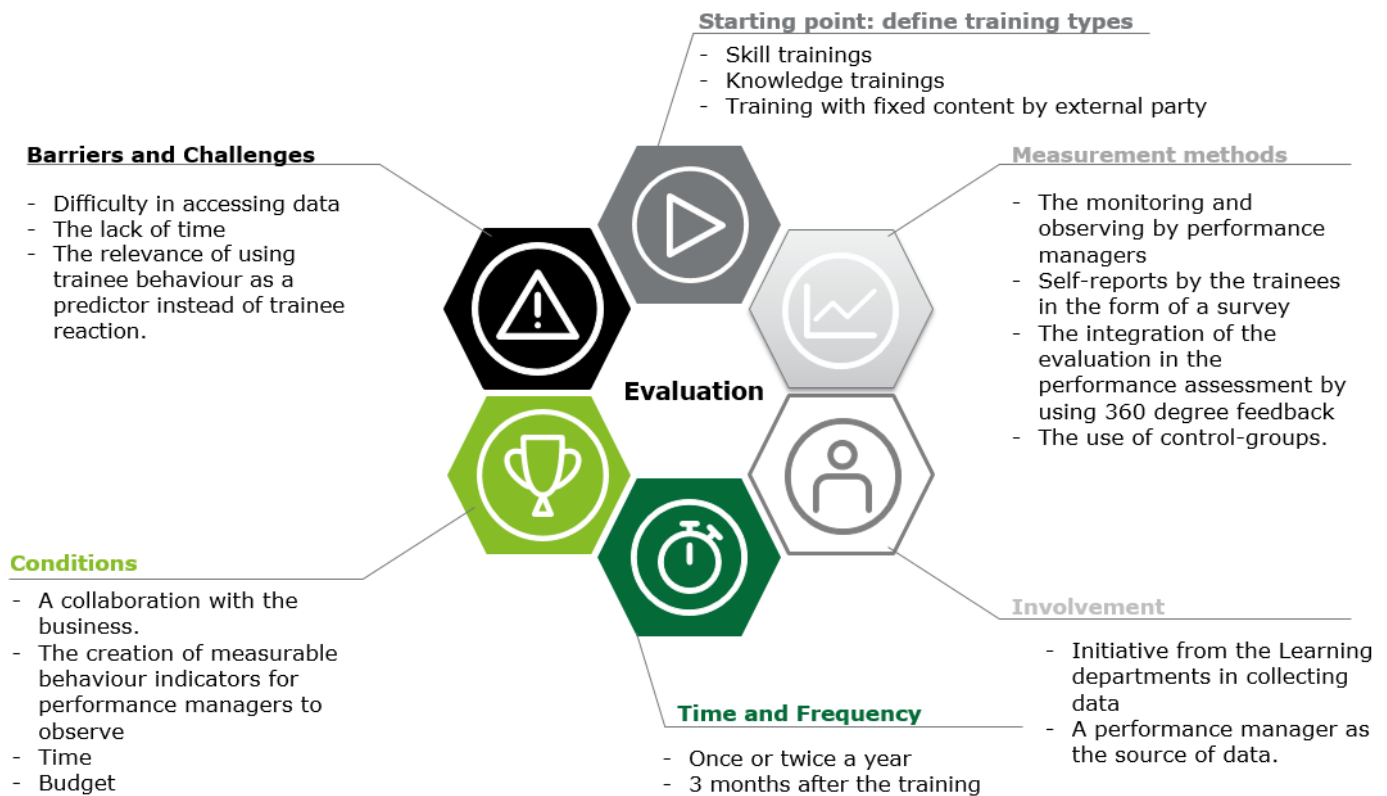
Starting point: define training types
- Skill trainings
- Knowledge trainings
- Training with fixed content by external party

Barriers and Challenges
- Difficulty in accessing data
- The lack of time
- The relevance of using trainee behaviour as a predictor instead of trainee reaction.

Measurement methods
- The monitoring and observing by performance managers
- Self-reports by the trainees in the form of a survey
- The integration of the evaluation in the performance assessment by using 360 degree feedback
- The use of control-groups.

Evaluation

Conditions
- A collaboration with the business.
- The creation of measurable behaviour indicators for performance managers to observe
- Time
- Budget

Involvement
- Initiative from the Learning departments in collecting data
- A performance manager as the source of data.

Time and Frequency
- Once or twice a year
- 3 months after the training

Figure 4. An overview of the evaluation characteristics and associated subcategories by learning professionals

6.1 Theoretical implications

Learning professionals in an organisation for professional services showed that there a lot of practical elements that should be taken into account in designing an evaluation approach that focuses specifically on the behavioural change in transfer performance. Existing evaluation frameworks, such as the IMTEE, do not include evaluation characteristics or evaluation criteria as an influencing variable. However, according to Velada et al. (2007) an evaluation after a training has a positive influence on the transfer performance. Therefore it is hypothesized that these evaluation characteristics are an influencing variable as well and could be integrated in the IMTEE. Subsequently, it is recommended to enrich the theory by confirming with quantitative research that the evaluation characteristics, that are identified and seen as relevant, are all part of the main construct 'evaluation characteristics'. Construct validity needs to be ensured. This could be followed by a research that investigates the relationship between the influencing variable evaluation characteristics and transfer performance. An implication for theory could then be that evaluation characteristics can be integrated in the evaluation frameworks, as is suggested by the researcher in figure 5.

6.2 Practical implications

This study gave Deloitte the opportunity to gain insights what evaluation characteristics are relevant and in designing an evaluation approach to determine training effectiveness on behaviour level. The outcomes of this study provided new input for elaborating existing evaluation frameworks regarding training effectiveness.

Evaluating training effectiveness presents opportunities for data-based decision making for a learning department. Therefore it is important for organisation that the design of an evaluation approach for training effectiveness is based on the one hand on literature and on the other hand on best practices. This study contributed to integrating the practical side of evaluating training effectiveness, where current literature focuses mostly on an overall theoretical view on evaluation. Practical questions such as how to measure, when to measure and what barriers and challenges arise in evaluating training effectiveness are not discussed within theory alone. Therefore this research tried to provide possible

leads (figure 4) for organisations to see what evaluation characteristics and concrete examples are available in designing an evaluation approach on behaviour level.
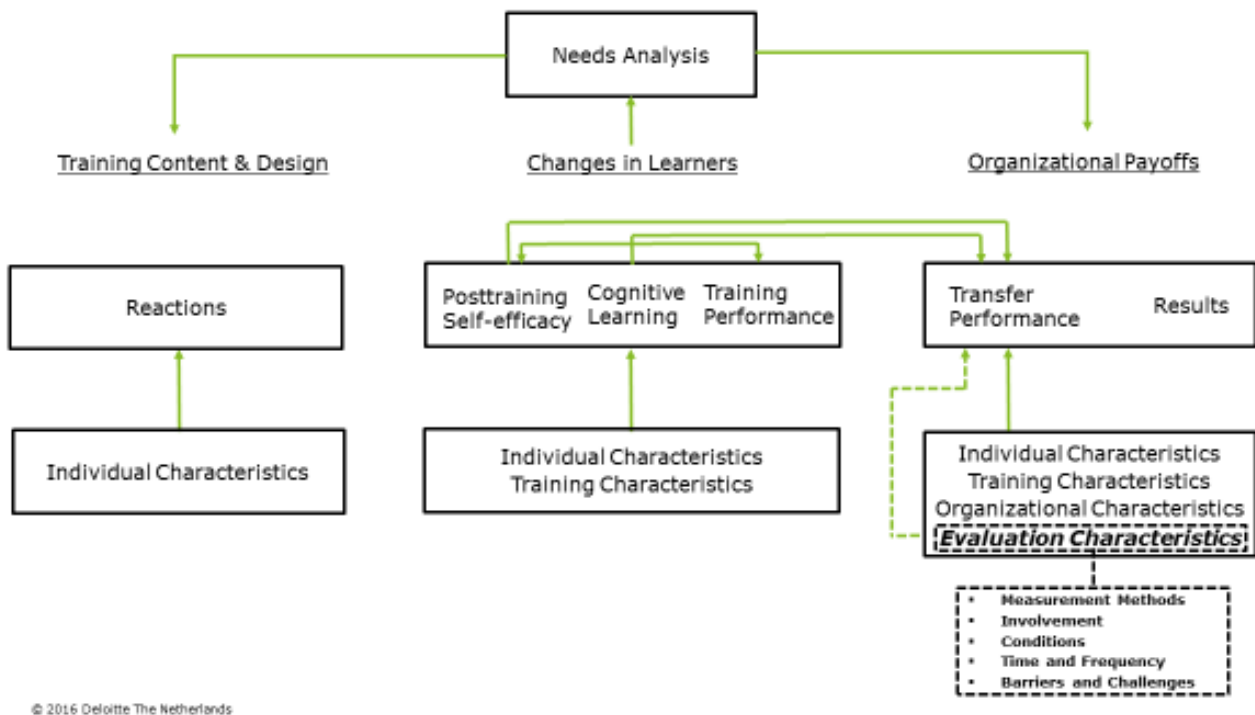

Figure 5. Elaborated Integrated Model of Training Evaluation and Effectiveness.

6.3   Recommendations for future research

The present study contributed to the existing literature about training effectiveness. However, this study was unable to encompass the entire field of effectiveness evaluation. The reader should bear in mind that the study is based on a pragmatic approach for organisations and also encounters the focus and practical questions that organisations might have. This case study is done in the context of Deloitte, which means that these results might not be generalizable to other organisations. However, it is assumed that organisations comparable to Deloitte might attach the same importance to the practical organisational characteristics and are representative for the sector of professional services. It is recommended to conduct further research in other industries and in different types of organizations to generalize these findings.

Next to that, the sample that is used contains 14 learning professionals. However, one of the results is that a collaboration with the business is required to conduct a proper evaluation. This study, due to the time limit, only investigated the perception of the learning professionals and excluded stakeholders within the business. A recommendation for further research is to explore what the perception of the business units is to test whether the approach suggested by literature and the training department corresponds. Case studies with an equivalent topic are necessary to generalize the results of this research.

Finally, it is recommended to confirm with quantitative research that the 5 categories, that are identified and seen as relevant, are all part of the construct 'evaluation characteristics'. This can be followed by a research that investigates the relationship between the influencing variable evaluation characteristics and transfer performance and the elaboration of the IMTEE or other evaluation frameworks.

This study has not only been beneficial in helping Deloitte to give insight in what evaluation characteristics are relevant in their field, but has given a contribution to future research as well. Future researchers have new input to integrate the practical side of evaluation in the existing frameworks. This future research is necessary to confirm or reject if the identified evaluation characteristics might influence transfer performance and play a role in an evaluation framework. This may hopefully result in the creation of a bridge between the theoretical and practical side of evaluation and make theoretical evaluation frameworks a guideline for the design of an evaluation approach on behaviour level in organisations.

References

Aggestam, L. (2006). Learning Organization or Knowledge Management- Which Came First, The Chicken or the Egg? *Information Technology and Control, 35*(3A), 295-302.

Aguinis, H., & Kraiger, K. (2009). Benefits of Training and Development for Individuals and Teams, Organizations, and Society. *Annual Revew of Psychology*, 451-474.

Alipour, M., & Salehi, M. (2009). A Study of on the Job Training Effectiveness: Empirical evidence of Iran. *International Journal of Business and Management, 4*(11).

Alvarez, K., Salas, E., & Garofano, C. (2004). An Integrated Model of Training Evaluation and Effectiveness. *Human Resource Development Review, 3*(4), 385-416.

AlYahya, M., & Norsiah, B. (2013). Evaluation of Effectiveness of Training and Development: The Kirkpatrick Model. *Asian Journal of Business and Management Sciences, 2*(11), 14-24.

Argyris, C. (1977). Organisational learning and management information systems. *Accounting, Organizations and Society, 2*, 113-123.

Argyris, C., & Schön, D. (1978). *Organisational Learning: A Theory of Action Perspective.* Reading, MA: Addison-Wesley Publishing Company .

Attia, A., & Honeycutt Jr, E. (2012). MEasuring sales training effectiveness at the behavior and results level using self- and supervisor evaluations. *Marketing Intelligence & Planning, 30*(3), 324-338.

Baldwin, T., & Ford, J. (1988). Transfer of Training: A review and directions for future research. *Personnel Psychology*, 63-105.

Bates, R. (2004). A critical analysis fo evaluation practice: the Kirkpatrick model and the principle of beneficence. *Evaluation and Program Planning, 27*, 341-347.

Bersin, J. (2015, January). *Predictions for 2015: Redesigning the Organizations for a Rapidly Changing World.* Retrieved from http://www.sap.com/bin/sapcom/en_us/downloadasset.2015-04-apr-17-19.predictions-for-2015-redesigning-the-organization-for-a-rapidly-changing-world-pdf.bypassReg.html

Blume, B., Ford, J., Baldwin, T., & Huang, J. (2010). Transfer of Training: A Meta Analytics Review. *Journal of Management, 36*(4), 1065-1105.

Boeije, H. (2008). *Analyseren in Kwalitatief Onderzoek: Denken en Doen.* Amsterdam: Boom Lemma .

Borman, W. (1991). Job behavior, performance and effectiveness. In M. Dunnette, & L. Hough, *Handbook of indistrial and organisational psychology* (pp. 271-326). Newbury Park, CA: Sage.

Brynjolfsson, E., & Hitt, L. (2011). Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance? *SSRN Electronic Journal*, 1-28.

Cheng, E., & Hampson, I. (2008). Transfer of training: A review and new insights. *International Journal of Management reviews, 10*(4), 327–341.

Cheng, E., & Ho, D. (2001). A review of transfer of training studies in the past decades. *Personnel review, 30*(1), 102-118.

Chiva, R. (2015). A Framework for Organisational Learning Types: Generative, Adaptive and Zero Learning. *Journal of Management and Organization, 21*(3), 1-19.

Crocetti, C. (2002). Corporate learning: A knowledge management perspective. *Internet and Higher Education, 4*, 271-285.

Donaldson, S., & Grant-Vallone, E. (2002). Understanding Self-report Bias in Organisational Behavior Research. *Journal of Business and Psychology, 17*(2), 245-260.

Feilzer, M. (2010). Doing mixed methods research pragmatically: Implications for the rediscovery of pragmatism as a research paradigm. *Journal of Mixed Methods Research, 4*, 6016.

Frye, A., & Hemmer, P. (2012). Program Evaluation models and Related Theories. *Medical Teacher*, 288-299.

Gielen, H., & Indesteege, L. (2004). *Rendement van leeractiviteiten.* Leuven/Den Haag: Acco.

Habicht, J., Victoria, C., & Vaughan, J. (1999). Evaluation designs for adequacy, plausibility and probability of public health programme performance and impact. *International Journal f Epidemiology, 28*, 10-18.

Hashim, J. (2001). Training evaluation: client's roles. *Journal of European indistrial Training, 25*(7), 374-379.

Hillsman, T., & Kupritz, V. (2011). A Conceptual Model of Training Transfer That includes the Physical Environment. *International Journal of Business Communication, 48*(2), 148-185 .

Holton, D., Bates, E., Burnett, M., & Carvalho, M. (1998). Factors Affecting Motivation to Transfer Training. *International Journal of Training and Development, 2*(1), 2-16.

Huber, G. (1991). Organisational Learning: The Contributing Process and Literatures. *Organisational Science, 2*(1), 88-115.

Jellema, F., Visscher, A., & Scheerens, J. (2006). Measuring change in work behaviour by means of multi-source feedback. *International Journal of Training and development, 10*(2), 121-139.

Johnson , D., Wang-Audia, W., & Krider, J. (2015). *Measuring Learning: Applying the Talent Analytics Framework to Learning.* Oakland: Bersin by Deloitte.

Kind, D. (1968). *Training within the Organization.* Cox & Wyman Ltd.

Kirkpatrick, D. (1994). *Evaualting Training Programs.* San Fransisco: Berret-Koehler.

Kirkpatrick, D., & Kirkpatrick, J. (2006). *Evaluating training programs* (3rd ed.). San Francisco: Berret- Koehler Publishers Inc.

Kraiger, K. (2002). Decision-based Evaluation. In K. Kraiger, *Improving training effectiveness in work organizations* (pp. 291-322). Mahwah, NJ: Lawrence Erlbaum.

Kraiger, K., Ford, K., & Salas, E. (1993). Application of Cognitive, Skill-Based, and Affective Theories of Learning Outcomes to New Methods of Training Evaluation. *Journal of Applied Psychology, 78*(2), 311-328.

Kraiger, K., McLinden, D., & Casper, W. (2004). Collaborative Planning for Training Impact. *Human Resource Management*.

Lance, C., Kavanagh, M., & Brink, K. (2002). Retraining climate as a predictor of retraining success and as a moderatr of the relationship between cross-job retraining time estimates and time to proficiency in the new job. *Group and Organization Management, 27*, 294-317.

Onwuegbuzie, A., Dickinson, W., Leech, N., & Zoran, A. (2009). A Qualitative Framework for Collecting and Analyzing Data in Focus Group Research. *International Journal of Qualitative Methods, 8*(3), 1-21.

Palys, T. (2008). Purposive Sampling. In L. Given , *The Sage Encyclopedia of Qualitative Research Methods* (pp. 697-698). Los Angeles: Sage Publications.

Pulichino, J. (2007). *Usage and Value of Kirkpatrick's Four Levels of Training Evaluation.* Doctoral dissertation.

Rajee, P., Madan, M., & Jayarajm, K. (2009). Revisiting Kirkpatrick's model - an evaluation of an academic training course. *Current Science, 9*(2), 272-276.

Roldán, J., & Leal, A. (2014). From Entrepeneurial Orientation and Learning Orientation to Business Performance: Analysing the Mediating Role of Organisational Learning and the Moderating Effects of Organisational Size. *British Journal of Management, 25*, 186-208.

Sachdeva, S. (2014). Effectiveness Evauation of Behavioural Training and Development Programmes. *The SIJ Transactions on Industrial, Financial & Business Management, 2*, 218-226.

Saks, A., & Burke, L. (2012). An investigation into the relationship between training evaluation and the quality and the transfer of training. *International Journal of Training and Development, 16*(2), 118-127.

Santos-Vijande, M., López-Sánchez, J., & Trespalacios, J. (2011). How organisational learning affects a firm's flexibility, competitive strategy and performance. *Journal of Business Research, 65*, 1079-1089.

Sessa, V., & London, M. (2011). Adaptive, generative and transformational learning in project teams. *Team Performance Management, 3/4*, 146-167.

Spitzer, D. (2005). Learning Effectiveness Measurement: A New Approach for Measuring and Managing Learning to Achieve Business Results. *Advances in Developing Human Resources, 7*(1), 55-70.

Tannenbaum, S., Cannon-Bowers, J., Salas, E., & Matthieu, J. (1993). *Factors that influence training effectiveness: A conceptual model and longitudinal analysis.* Orlando, FL: Naval Training Systems Center.

Tauber, T., & Wang-Audia, W. (2015). *The State of Learning Measurement* . Oakland: Bersin by Deloitte.

Tracey, J., Hinkin, T., Tannenbaum, S., & Mathieu, J. (2001). The Influence of Individual Characteristics and Work Environment on Varying Levels of Training Outcomes. *Human Resources Development Quarterly, 12*, 5-23.

Tracey, J., Tannenbaum, S., & Kavanagh, M. (1995). Applying Trained Skills on the Job: The Importance of the Work Environment. *Journal of Appied Psychology, 80*(2), 239-252.

Velada, R., Caetano, A., Michel, J., Lyons, B., & Kavanagh, M. (2007). The effects of training design, individual characteristics and work environment on transfer of training. *International Journal of Training and Development*, 282-294.

Weisweiller, S., Nikitopoulos, A., Netzel, J., & Frey, D. (2013). Gaining insight to transfer of trainng through the lens of social psychologu. *Educational Research Review, 8*, 14-27.

Wick, C., Pollock, R., & Jefferson, A. (2010). *The Six Disciplines of Breakthrough Learning.* San Fransisco: Pfeiffer.

Yin, R. (2003). *Case Study Research: Design and Methods.* Thousand Oaks: Sage Publications.

Appendices

Appendix A   Coding scheme

Main category	Sub category	Code	Example
Training types	Skill training	T1	3.295 M Het gaat voornamelijk om skills bedenk ik mij net.
	Knowledge training	T2	3.2 U Ja een groot deel is vaktechnisch, we hebben de Summerschool bij Tax. Dat is eigenlijk 100 procent vaktechnisch. En daarnaast hebben we wat skilltrainingen, bijvoorbeeld de New-hire week, dat is met name op skills gericht. Bij de Summerschool zit voor de analisten ook nog een klein beetje skills. Dat zijn denk ik wel de twee poten waarop wij zitten. De echt pure vak techniek voor Tax en de skills.
	Behavioural training	T3	3.291 J Programma's waarbij je binnen een korte tijd een gedragsverandering wil bewerkstelligen. Misschien is dat het wel.
	Other training	T4	1.44 J Ja, het is een soort methodologie maar dan gegoten in een template.
Involvement	Trainee	I1	3.52 M Bij de persoon zelf. En je zou hem inderdaad, je zou hem als je echt wil kijken naar wat de training heeft opgeleverd en je wilt kijken inderdaad naar behaviour, dan moet je hem daarna evalueren.
	Supervisor	I2	1.525 J Ja of dat je het juist aan je performancemanager vraagt. Dat die..
	Peer	I3	3.133 S Naast een performancemanager, wie heeft er nog meer zicht op een gedragsverandering? 3.134 U Peers denk ik ook
	Client	I4	2.28 AB Ik zou hem heel erg nog willen doortrekken maar dat is misschien nog wel meer vanuit mijn rol omdat ik erg op de klant gericht bent, de externe klant. Uiteindelijk moet je bij de eindgebruiker gaan informeren van: is die kennis inderdaad goed, voldoende en up to date? Nouja, dan zou je eigenlijk nog een stap verder moeten gaan.
	Learning department	I5	1.96 J Nouja, het is interessant dat we daar kennis van nemen maar wij hoeven niet, Learning hoeft niet die rol te hebben om het maar.. Wij zijn geen politieagent zou ik maar zeggen of inspecteur, maar wel.
	Other involvement	I6	3.110 M Talent meer denk ik hoor
Measurement methods	Observation	M1	3.67 MM En je hebt het geleerde dan ook nog in de praktijk gebracht tijdens je onderhandelingsgesprekken en dus er moet iemands naast zitten die weet wat de inhoud is van de training en hebben gezien van: hee dat heb jij in de praktijk gehad? Dus dat is best..
	Interview	M2	1.402 R Ik denk dat het heel waardevol zou zijn als we van iedere function 2 mensen hebben in de pre-manager journey die bijvoorbeeld gaan kijken van: hoe pak je dit nou aan? Waar loop je tegenaan met het kiezen van je trainingen, dat soort dingetjes?
	Site visit	M3	2.120 AH Mystery shopper ja. Dat soort dingen, dat is in onze context natuurlijk moeilijker maar uiteindelijk wil je wel vanuit die invalshoeken bekijken hoe die professional functioneert. En ik denk dat we vanuit Learning een rol hebben om dat te organiseren maar uiteindelijk is de business verantwoordelijk dat het probleem wordt opgelost. Want zij lopen ook tegen het probleem aan. Wij niet. Het maakt mij niet uit of dat probleem wordt opgelost of niet. Ik heb er geen last van.
	Self-report	M4	2.67 C Het ligt er ook aan hoe. Als je dan vervolgens een ding moet gaan invullen wat je een kwartier de tijd kost. Maar ja als je gewoon even

			kort krijgt: het is nu drie maanden geleden. Om de effectiviteit te meten voor jezelf maar ook voor ons zal het.. en weet ik veel dat je dan 3 vragen hebt van: merk je dat je gedrag is veranderd? Ja/nee en in welk opzicht. Ik denk dat als je het kort houdt en de mensen, de deelnemers gewoon meeneemt in het belang hiervan. Want ik vind, wat jij ook zegt: het is onwijs interessant om dit te weten gewoon als deelnemer zelf, maar ook voor het bedrijf en zo ook weer voor de klant. Het is een investering ook die Deloitte in jou doet als medewerker, dat je mag deelnemen. Dan vind ik dat je ook wel iets terug mag vragen.
	Peer report	M5	2.82 AH Nou , dat je na afloop van een training ook een soort APR uitstuurt naar je directe collega's met: dit was de doelstelling van de training die ik heb gevolgd de afgelopen maanden en ik heb aan dit en dit gedrag gewerkt. Kun je me feedback geven op wat je me hebt zien veranderen in gedrag. Of heb ik dingen anders aangepakt dan voorheen?
	Control group	M6	1.426 An Nouja, als je bijvoorbeeld een teamtraining doet dan kan je natuurlijk ook gewoon zo'n team laten evalueren. 1.427 J En dan bijvoorbeeld afzetten tegen een vergelijkbaar team?
	Organisational audit	M7	1.79 J Nouja, ik zit even te denken, binnen Audit heb je natuurlijk een soort.. je hebt een controlemechanischme van practice reviews. Dus daar wordt gewoon periodiek gekeken, steeksproefsgewijs gekeken naar: hoe worden de audits gedaan? Die worden gewoon nagelopen op gewoon allerlei punten gecontroleerd door de interne afdeling zeg maar. Die kijken volgens mij wel naar dit soort elementen van: zijn de juiste checklists gebruikt, is het team op de juiste manier bemand, hebben mensen de juiste accreditaties?
	Performance analysis	M8	2.83 C Wat we in het begin al zeiden: dat je het een onderdeel maakt van de performcyclus. Ik ga deze training volgen want ik wil… Ik heb als target dat ik aan dit gedrag wil werken en dat je dan met je mid-year op kunt halen van .. dat je het niet via iValue doet bij wijze van maar gewoon specifiek op dit gedrag feedback gaat vragen aan collega's, maar dat kunnen ook gewoon opdrachtgevers zijn.
	Expert review	M9	No example in interview
	Supervisor	M10	1.216 L Ja maar ik denk niet dat je misschien niet eens de deelnemers zou moeten.. ik denk dat je de managers of de performance managers of whoever, of de talentportefeuillehouder.
	Other measurement method	M11	3.25 C Heel veel van die vaktechnische overleggen en daar leren ze toch ook heel van 3.26 M Daar leren ze heel veel 3.27 C Daar moeten ze toch demonstreren en laten zien wat ze kunnen maar dat is niet Learning gerelateerd. Althans, wij houden dat niet bij of kunnen daar op 1 of andere manier.. of wel?
	Survey	M12	1.226 J Dat zou persoonlijk kunnen maar dat zou ook wel middels een survey kunnen. Dat je ze gewoon vraagt: Kun je eens kijken wat je.. he? Of je verbetering ziet op deze punten. Of in hoeverre..
	Behaviour indicators	M13	3.72 M Ik weet niet of dat kan. Dus ik denk dat je concreet moet zijn en dat je het heel klein moet hebben 3.73 C En ik denk ook niet, wat jij zegt, die concreetheid in van: ja ik vind dus dat hij het goed doet..
Barriers	Difficulty accessing data	B1	2.199 C En dat is nou net een hele lastige om te meten want je moet maar net weet je, alle analisten worden daardoor heen gehaald. Nou ja, sommigen komen dit pas over een jaar tegen en anderen gebruiken het misschien onbewust.
	No management-support	B2	3.118 MM Je staat gewoon onder aan het lijstje want ze zijn zo ontzettend druk dat dit er gewoon niet bij kan.

	Time-consuming	B3	2.10 M Ja je vraagt best veel tijd van mensen. Want of het moet via zelfreflectie of je moet een performancemanager vragen of zij waarneembaar andere dingen zien. Of je moet zelf gaan observeren. Dus op een of andere manier kost dat toch tijd en het is ook wel weer een leermoment natuurlijk dat je nog een keer reflecteert, maar als je ziet hoe weinig tijd mensen eigenlijk hebben voor mede leermomenten.
	Not urgent for training department		2.23 AH Ja de vraag is of je dat dan moet weten want als het alleen maar als doel heeft om de kennis aan te leren over hoe iets in elkaar zit dan is het alleen maar denk ik voor ons belangrijk om te meten of die kennis daadwerkelijk ook beklijft en of diegene het ook heeft begrepen wat er in die training is verteld maar niet per se of die daar dan ook iets mee doet.
	Expensive	B5	Not mentioned in the interview
	Not enough expertise	B6	1573 L Maar ik denk ook dat we dat best de business zelf kunnen vragen. Of de business whoever, hoe ze dat zelf willen meten? Want soms hebben wij helemaal niet de kennis of de tooling of ..
	Reaction evaluation is enough	B7	2.204 C Eerlijk, minder. Ik wil wel weten of ze het een goede training vonden, of het goed gefaciliteerd is, of de manier, daar zijn we vanuit Learning heel veel mee bezig. Weet je, hebben ze.. en dan kun je vanuit Learning ook erop inspelen. Is een facilitator bijvoorbeeld mega slecht beoordeeld, nouja..
	Other barriers	B8	3.256 J Maar nu beschrijf jij wel de ideale situatie en misschien gaat het zo binnen Tax he dat het bij jou binnenkomt en dat het via D&D wordt besproken en zo verder. Maar het gebeurt ook heel vaak via .. dat het via andere kanalen gaat en dat het al bijna is geregeld en dat alleen eventmanagement er nog wat mee kan doen. Dus het is ook heel vaak dat dus die vraag niet zo gestructureerd binnenkomt waardoor ook die afstemming daardoor wel lastiger wordt.
Influencing factors	Organisational culture	F1	Not mentioned in the interview
	Organisational climate	F2	Not mentioned in the interview
	Supervisor support	F3	Not mentioned in the interview
	Other influencing factors	F4	1.369 J Want jij vroeg ook net: ja wat doe je dan? Ik heb namelijk ook heel vaak gemerkt dat het komt omdat mensen gewoon de tijd er niet meer voor nemen, ze denken dat de training een soort magisch moment is waarin je, waarna je niks meer hoeft te doen en je lekker over kunt gaan tot de orde van de dag. En dan blijkt na een maand dat je er eigenlijk niet zoveel van gebakken hebt en dan.. snap je? Dus de vraag is een beetje van: onze invloed is beperkt. We kunnen wel onze invloed vanuit.. We zouden vanuit Learning kunnen zeggen: we proberen daar verder en verder in te gaan, of heel ver in te gaan, maar het zit.. dat die transfer er niet is zit du svaak ook wel aan de kant dat de manager niet aangehaakt is, niet ernaar vraagt van: joh wat heb je er mee gedaan? Of dat iemand er inderdaad gewoon de tijd niet voor neemt.
Conditions	Conditions	C1	
Time and frequency	Time and frequency	TF1	
Not relevant	Not relevant	NR1	

Appendix B   Survey demographical data

1. Wat is uw geslacht?

◯ Man

◯ Vrouw

2. Wat is uw leeftijd?

[   ]

3. In welke regio bent u woonachtig?

◯ Rotterdam en/of randgemeente(n)

◯ Amsterdam en/ of randgemeente(n)

◯ Den Haag en/of randgemeente(n)

◯ Utrecht en/of randgemeente(n)

◯ Anders

4. Wat is uw hoogst voltooide opleiding?

◯ Basisonderwijs

◯ Lager/voorbereidend beroepsonderwijs (lbo/vmbo)

◯ Middelbaar algemeen voortgezet onderwijs (mavo)

◯ Middelbaar beroepsonderwijs (mbo)

◯ Hoger algemeen voortgezet onderwijs (havo)

◯ Voorbereidend wetenschappelijk onderwijs (vwo)

◯ Hoger beroepsonderwijs (hbo)

◯ Wetenschappelijk onderwijs (wo)

◯ Anders

5. Wat is uw functie binnen de afdeling?

○ Director

○ Learning advisor

○ Programmamanager

○ Eventmanager

○ Design and Development

○ Anders

6. Hoelang bent u werkzaam bij Deloitte?

○ Minder dan een jaar

○ 1 tot 3 jaar

○ 3 tot 5 jaar

○ 5 tot 10 jaar

○ 10 tot 15 jaar

○ 15 tot 20 jaar

○ 20 tot 25 jaar

○ Langer dan 25 jaar

7. Hoelang werkt u binnen het gebied van Learning en Development?

○ Minder dan een jaar

○ 1 tot 3 jaar

○ 3 tot 5 jaar

○ 5 tot 10 jaar

○ 10 tot 15 jaar

○ 15 tot 20 jaar

○ 20 tot 25 jaar

○ Langer dan 25 jaar