UNIVERSITÄT PADERBORN
Die Universität der Informationsgesellschaft

FACULTY OF BUSINESS ADMINISTRATION AND ECONOMICS

Working Paper Series

Working Paper No. 2015-05

Advances in Training Evaluation - Psychological, Educational, Economic, and Econometric Perspectives on the Kirkpatrick Model

Christiane Hinerasky

December 2014

Advances in Training Evaluation –

Psychological, Educational, Economic, and Econometric

Perspectives on the Kirkpatrick Model

Christiane Hinerasky[1]

University of Paderborn

Abstract

Research on the process of training evaluation has progressed in many independent fields. In our study, we combine the unique views on training, and training evaluation from the fields of psychology, education, economics, and econometrics. Psychology and education provide knowledge on how to conduct training, and they emphasize important individual and environmental factors that may facilitate the transfer of skills and prevent skill decay. However, empirical methods offer a sound way of not only testing the underlying theoretical hypotheses, such as Human Capital theory in economics or psychological cognitive theories, but also quantifying individual and operational performance effects. With this multidisciplinary approach, we are able to rethink existing views on the underlying learning mechanisms and generate new insights into this complex and multifaceted economic subject of returns to training.

JEL Classification: M53, J24, C31

Keywords: Training, Evaluation, Research Review, Multidisciplinary Approach, Reaction to Training, Learning, Transfer, Empirical Methods, Selection, Selection on Unobservables, Dynamic Evaluation Concepts

[1] Department of Management, University of Paderborn, Warburger Str. 100, D-33098 Paderborn, Germany.
Email: christiane.hinerasky@wiwi.uni-paderborn.de, phone: +49-(0)5251-60-3692, fax: +49-(0)5251-60-3550

1 Introduction

The first systematic approach for the evaluation of training was introduced by Kirkpatrick, who evaluated the outcome and success of training based on four distinct criteria: reaction, learning, behavior, and results (Kirkpatrick, 1979). His four-level taxonomy has become popular in business and academia, as it addresses the need to understand training evaluation "simply yet systematically" (Alliger et al., 1997, p. 342). Kirkpatrick's evaluation taxonomy has been extended and complemented ever since. Scholarly interest and the importance of training in work organizations are reflected by the regular publication of thematic reviews in the *Annual Review of Psychology* since 1971, which addresses training benefits not only for individuals but also teams, organizations, and the society (Arguinis and Kraiger, 2009; Campbell, 1971; Goldstein, 1980; Latham, 1988; Salas and Cannon-Bowers, 2001; Tannenbaum and Yukl, 1992; Wexley, 1984). Whereas the hands-on evaluation within firms has not advanced much due to the lacking knowledge among HR managers, economic and quantitative approaches to the evaluation of training have emerged, and they now combine various streams of literature from labor economics, educational research, psychology, microeconometrics, and statistics, each adding its own valuable insight.

Labor Economists build upon Becker's Human Capital theory (Becker, 1993, 1964) in order to assess how training participation can increase individuals' productivity. Human Capital theory, however, does not provide information on *how* an increase in productivity actually arises, i.e. how learning affects productivity. This gap can be filled by educational research, which offers a vast amount of literature on the transfer of learning among pupils, which should be used to provide insight into the transfer of training. When it comes to quantitatively measurable outcomes of training, we can build on the evaluation research of labor market programs in applied microeconometrics, which deals extensively with the fundamental evaluation problem and selection bias, as the identification of causal effects beyond the measurements of mere correlations is the key to empirical research in the social sciences (Altonji, Elder, and Taber, 2005).

This paper attempts to outline, based on Kirkpatrick's four-level taxonomy, how these streams of literature complement each other and offer new conceptual perspectives are on the different levels of training evaluation. The following summary makes no claim of being complete, as the theories and research discussed are selective and illustrative. But we believe this multidisciplinary approach is useful due to the increasing fragmentation of knowledge generated by researchers in various training subfields (Arguinis and Kraiger, 2009). By combining these streams of literature, existing views on the underlying learning mechanisms can be extended and reconsidered, and new insights into this complex and multifaceted economic topic can be generated. Based on existing psychological and managerial approaches for the provision of training, this research shows how the economic program evaluation adds to the evaluation of training provisions, along with determining which econometric and statistical tools offer new perspectives on the interpretation of results. This will lead to a better understanding of the factors that impact learning and its transfer process and, in turn, can fill in the gaps in the underlying theoretical considerations.

The paper proceeds as follows: First, we summarize the managerial and psychological views on Kirkpatrick's evaluative steps of reaction, learning, and transfer, and provide econometric insights regarding the results level. Further, there is a discussion on the new statistical methods that can help tackle the specific features present in company training data. The following section briefly discusses conclusions that can be drawn regarding the ROI calculations for training programs, and the last section concludes this paper.

2 Reaction to training

Whether training is used to improve current job skills, to prepare for career advancement, to retool for new or changing job requirements, or to enter into an organization, Kirkpatrick's first training evaluation level is based on the *reaction* of trainees to training; it focuses on how much the trainees liked the program and/or how well it has been accepted (Kirkpatrick, 1979). Presently, most HR managers still base their training evaluation on so-called "happy sheets," which are simple surveys of these reaction criteria (Alliger et al., 1997). Well-designed and well-administered training programs are expected to result in positive trainee reactions, learning, behavior changes, and improvements in job-related outcomes. Cognitive and social psychology seeks to determine how these criteria interrelate and how the heterogeneity of training outcomes among participants within the same program may be explained.

For illustrative purposes, we present the concepts of trainability and training fulfillment, which derive from the social psychologies and relate trainee reactions to further desired post-training behavior. The concept of training fulfillment refers to "the extent to which training meets or fulfills a trainee's expectations and desires" (Tannenbaum et al., 1991, p. 759) and relates high fulfillment rates, by implication, to the development of desired post-training attitudes. Based on the fact that unmet expectations about professional life have been associated with dissatisfaction and higher turnover, Tannenbaum et al. (1991) "investigate how trainees' expectations and desires before training, and their subsequent perceptions of what occurred during training, can influence the development of post-training commitment, self-efficacy, and motivation" (p. 759). Since only 16 % of the variance in trainee performance can be attributed to ability (Robertson and Downs, 1979), individual differences in the attitudes and expectations must be central influences on training effectiveness (Tannenbaum et al., 1991). In earlier research, Maier (1973) and Noe and Schmitt, p. (1986, p. 498) implied that low motivation will lead to poor after-training performance, "even if trainees possess the prerequisite skills needed to learn the training program content". Therefore, Noe and Schmitt (1986) integrated motivation and environmental influences into their concept of trainability, which should explain why learning, behavior changes, and performance improvement are different among training program participants. They concluded that satisfaction with the administration and the content of the training program leads to greater training fulfillment and higher motivation after training, and those positive impressions will be carried into the workplace (Noe and Schmitt, 1986). Similarly, Leach and Liu (2003) found that trainees who had positive reactions to sales trainings were more likely to learn the material, and trainees with higher levels of knowledge retention were more likely to apply the material in the work environment.

Given the commonness of trainee reaction measures in the workplace and in research, attitude theorists further developed reaction measures by recognizing the difference between affective reactions and more behaviorally evaluative responses such as utility judgment (Alliger et al., 1997). Warr and Bunce (1995) added the difficulty of training as an additional measure to Alliger et al.'s (1997) affective and utility measures of training programs; in hierarchical regression analysis, they found affective reactions to be unrelated to other desired training outcomes. Alliger et al. (1997) summarized the findings by stating that "liking does not equate to learning or to performing" (p. 353), but utility-oriented questions may be able to predict or indicate learning transfer. Finally, research on mood and emotions was used by Brown (2005) to extend the theoretical framework on reactions. Similarly, the results suggested that enjoyment and relevance should be seen as hierarchical facets of overall training satisfaction, and trainee reactions are mainly influenced by trainee characteristics. In a meta-analysis on reaction evaluation, Sitzmann et al. (2008) confirmed these findings in the sense that trainee characteristics have a moderate effect on reactions.[2]

Despite years of advice from economic researchers stating that trainee reactions provide training evaluation information that is of very limited use, reactions of trainees remain the most commonly used measures of training effectiveness (Long, DuBois, and Faley, 2008). In fact, many firms still rely on self-administered reports that are completed by the trainee to evaluate training courses (Aragon and Valle, 2013). However, simple characteristics of the training course are primarily captured by reaction measures. These are, only to some extent, able to predict changes in motivation from pre- to post-training. Rowold (2007) ran regression analyses and found that only expectation fulfillment significantly impacted knowledge acquisition (as opposed to pre-training expectations or post-training reaction). Hence, satisfaction with training seldom causes learning success, let alone transfer success (Arthur et al., 2003; Colquitt, LePine, and Noe, 2000). However, the main driving force for transfer motivation and transfer itself is (perceived) practical relevance, as opposed to trainee satisfaction (Liebermann and Hoffmann, 2008). All models agree, however, that irrespective of whether positive reactions to training are predictors for learning, negative reactions to training often have adverse effects.

[2] This relationship was already proposed by Cannon-Bowers et al. in 1995, along with other models of training effectiveness (see Alvarez, Salas, and Garofano, 2004 for a relationship theory between trainees' characteristics and subsequent learning).

3 Learning, knowledge, and skill acquisition

Kirkpatrick's second level focuses on *learning*, i.e. the level of knowledge and skill acquisition. This level asks what principles, facts, and techniques were understood and absorbed (Kirkpatrick, 1979). Learning constructs are derived from a variety of research domains, such as cognitive, social, and instructional psychology and human factors (Kraiger, Ford, and Salas, 1993). In the most recent decades, the training and psychology literature both have heavily discussed learning models and influential factors on learning. While research in psychology delivers theoretical models on how learning may occur, the training literature and research in education attempt to identify influencing and moderating effects on the individual and environmental levels that may foster or hinder actual learning.

Tannenbaum, Cannon-Bowers, and Mathieu's (1993) integrative framework states pre-training and during-training conditions that may influence skill acquisition based on individual characteristics (cognitive ability, self-efficacy, goal orientation, training motivation) and environmental characteristics (training induction and pre-training environment)[3]. Similarly, Kraiger, Ford, and Salas (1993) derive a multidimensional learning model, which uses more precise measures of learning outcomes (also see Salas et al., 2012) and was reviewed in more than 125 studies (Ford, Kraiger, and Merritt, 2010). A major influence on actual learning is seen in the motivation to learn, which refers to a condition when "trainees believe that training is relevant and are willing to exert effort in the training environment" (Salas et al., 2012, p. 79). This "motivation to learn" is again influenced by both individual characteristics (self-efficacy, mastery orientation, prior successful training participation) and organizational influences (training support) (for a review, see Salas et al., 2012 or Colquitt, LePine, and Noe, 2000). Using 20 years of research on training motivation, Colquitt, LePine, and Noe (2000) based their theory of training motivation on a meta-analytic path analysis and found support for an integrative model of training motivation, where the influences of individual and situational characteristics on skill acquisition are partially mediated by self-efficacy, valence, job variables, and motivation to learn. Special interest is given to the examination of feedback and feedback specificity on the individual

[3] also see Cannon-Bowers et al. (1995) and Salas and Cannon-Bowers (2001).

learning process and how it can enhance or inhibit learning (Goodman and Wood, 2004; Goodman, Wood, and Chen, 2011).

To date, the "training field has grown exponentially" (Arguinis and Kraiger, 2009, p. 452), not only in the psychological literature, but also in related fields such as human resource management, instructional design, human resource development, human factors, and knowledge management. The management literature analyzes learning based on the adoption of organizational behavior theories, including goal setting, social learning theory, expectant theory, social exchange theory and justice theory (see Locke et al., 1984; Gist, Stevens, and Bavetta, 1991; Mathieu, Tannenbaum, and Salas, 1992; Quiñones, 1995). Insights from the psychological literature are gathered by Noe, Tews, and McConnell Dachner (2010), which adds to the management literature by providing a better understanding of learner motivation, subsequent knowledge and skill acquisition, and workplace learning.

To measure the actual amount of learning, traditional tests of declarative knowledge are used to index learning. This process is relatively common in the training literature (but not in economics). Kraiger, Ford, and Salas (1993) reviewed the theoretical work in psychology and distinguished three learning outcomes: cognitive, skill-based, and affective learning. Although learning may occur in each of these three domains, they each need different procedures or measures to infer learning (Kraiger, Salas, and Cannon-Bowers, 1995). Alliger et al. (1997) distinguished between immediate post-training knowledge and knowledge retention to model skill decay. In fewer cases, the assessment of (new) knowledge is measured at a later time, rather than or in addition to being measured immediately after training participation. In economics, such measures from multiple choice test responses, answers to open-ended questions, listings of facts, and so forth are seldom derived. To our knowledge, there is only one study, by Hinerasky and Fahr (2014a), that explicitly measures the amount of learning from multiple choice tests at the end of training and uses this information to evaluate the training program. More often, companies rely on participants' self-assessments on whether and how much knowledge has increased. LeRouzie, Ouchi, and Zhou (1999) investigated whether participants' statements on what they have learned can be a valid proxy of the actual amount of learning. In a non-representative analysis at the World Bank Institute, self-assessment was found to be positively correlated with the amount of actual learning. However, the correlation was too weak to rely on self-reported data to measure actual learning. Similarly, Dixon (1990) used post-test results of 1200 employees of a large manufacturing company and found no relationship between participants' perceptions of how much they learned and their actual test scores. Therefore, self-reported data cannot be relied upon in order to measure knowledge acquisition.

Since skill decay has been identified as a major problem in training (Salas et al., 2012), with trainees losing over 90% of knowledge in the first year after training participation (Arthur et al., 1998), the question of when to measure skill acquisition is closely related to learning itself. Intuitively, knowledge assessment should occur immediately after training, and there should be a timely gap for the possible identification of skill decay, as the overall retention decreases with longer periods of non-use or missing practice (Arthur et al., 1998). Blume et al. (2010) also found that post-training measures of knowledge and transfer decrease by 62.5% (from .48 to .18), in studies where knowledge is assessed with a time lag, as opposed to studies without.

4 Changing behavior and using training on the job

The transfer of training into the workplace is a "black box" in the economic literature (DeGrip and Sauermann, 2013). Human Capital theory cannot explain how learning actually happens. Rather, it assumes that larger amounts of skill inherently lead to higher productivity, which can then be measured by higher wages. The mechanisms by which training affects performance, e.g. by larger amounts of skills or higher motivation, remains unclear. A study by Hinerasky and Fahr (2014b) found performance effects of training to unfold during training rather than after training, and they concluded that such non-persistent treatment effects may not reflect a human capital increase; it remains unclear whether (non-persistent) human capital has been built up or whether knowledge had not been permanently transferred into the workplace. In view of the companies' high expenditures on training workplace transfer is considered to be very important for investment success. To gain insight in the transfer mechanisms, an integrated research perspective is needed (Segers and Gegenfurtner, 2013). We will use this integrated perspective to shed some light on Kirkpatrick's third evaluation level, building on a combination of various streams of literature on the transfer of training from education, psychology, sociology, and economic standpoints.

Kirkpatrick's third level, *behavior*, examines on-the-job use of the learned principles, facts, and techniques (Kirkpatrick, 1979). Transfer has been defined as either the "use of a trained skill" or "the effectiveness in performing the trained skill" in the workplace (Blume et al., 2010, p. 1070). Again, the psychological literature delivers instruments for measuring transfer success given their models of transfer of training (e.g. Grossman and Salas, 2011). The management literature also delivers empirical evidence on influencing factors on training transfer (e.g. Burke and Hutchins, 2007; Blume et al., 2010). Educational research should also be consulted, since insight on the transfer of training may be derived from literature regarding the transfer of learning (Volet, 2013).

Exactly how transfer and changes in work performance occur, however, is a complex and multifaceted psychological phenomenon (Volet, 2013). The psychological literature formulates in fairly general terms, that "factors before, during, and after training can influence the extent of transfer to the job" (Salas et al., 2012, p. 88). In an effort to explain the underlying mechanisms of training transfer, Weisweiler et al. (2013) focused on concepts from social psychology and combined them with well-known concepts in training transfer research, in order to determine why certain people do not change their behaviors after training, even if they liked training and learned a great deal. These individual differences in the transfer of training are explained by personal and environmental factors. Similarly, Holton (2005) and Holton, Bates, and Ruona (2000) examined transfer characteristics and organized them in four groups: trainee characteristics (learner readiness and self-efficacy), trainee motivation (motivation to transfer, transfer effort to performance expectations, and performance to outcome expectations), work environment (performance coaching, supervisor support, supervisor sanctions, peer support, resistance-openness to change, positive personal outcomes, and negative personal outcomes), and ability (perceived content validity, personal capacity for transfer, transfer design, and opportunity to use). While researchers have identified single variables and their meaning for the transfer process, Blume et al. (2010) reviewed the literature by means of a meta-analysis to shed light on the dynamics inherent in the transfer process. On an individual level, the transfer of training is mostly influenced by cognitive ability, then by conscientiousness and voluntary participation, then by a supportive work environment on the firm level, followed closely by peer and supervisory support.[4] A supportive post-training environment (positive transfer climate) affects employees' mindsets, which in turn will determine whether they use what they have learned in training (Blume et al., 2010; Rouiller and Goldstein, 1993). As employees may (or may not) elect to use the acquired knowledge once they return to "life as usual" (Perkins and Salomon, 2012), most attention in answering the transfer question was paid to the individual rather than the organizational level. On an individual level, the question of learning and transfer is rather cognitive (Volet, 2013). However, estimations show that only 10% to 15% of learned material can be transferred to the workplace (Kauffeld, Lorenzo, and Weisweiler, 2012). As cognitive ability cannot be influenced, the work environment has the most potential for improving the transfer situation.[5] Transfer environments are not only unique to each organization, but also to each training application. Since only 7 % to 9 % of skill acquisition in organizations comes from formal training, trainees must continue to learn on the job (Tannenbaum, 1997). Therefore, opportunities to practice are related to a supportive post-training environment.

Eventually, individual level outcomes should be assumed to eventually influence higher-level outcomes, which is crucial to training effectiveness. So far, research on the transfer of training and on organizational-level outcomes has progressed in isolation of each other. The link between both streams of literature is seen in the vertical transfer of training, which refers to the "upward propagation of individual-level training outcomes that emerge as team- and organizational-level outcomes" (Salas and Cannon-Bowers, 2001, p. 474). Just recently, Saks and Burke-Smalley (2014) attempted to bridge the gap between the research on transfer of training and firm performance. The measurement of organizational-level results, which is necessary to explore this question, will be extensively discussed in the next section, with the help of the economic program evaluation.

[4] Effect sizes between transfer and influential factors are heavily influenced by the source of the transfer measure. When the predictor variables and transfer are both measured simultaneously by the trainee, as opposed to an objective measure by others, same-source and same-measurement-context (SS/SMC) bias consistently inflated the relationships between the examined constructs.
[5] It was shown that motivation to transfer can predict transfer success. Motivation to transfer, however, is majorly influenced by the work environment.

5 Changes in behavior transform into tangible results

Kirkpatrick's fourth and last level, *results*, focuses on the improvements in tangible individual or organizational outcomes (Kirkpatrick, 1979). Outcomes may be found in three different categories: learning, individual performance, and organizational results (Holton, 1996). In a meta-analysis of the relations among Kirkpatrick's training criteria, Alliger et al. (1997) failed to include the level of results, as only very few studies dealt with the effect of training on individual performance or organizational outcomes. Kozlowski et al. (2000) also concluded that there was a gap in the training literature. Even though "a goal of training is to enhance organizational effectiveness, the models, methods, and tools of training focus on the individual level" (Tharenou, Saks, and Moore, 2007, p. 253). Tharenou, Saks, and Moore (2007) were the first to conduct a meta-analysis on the relation of training and organizational-level outcomes based on research in the management literature. Again, they concluded that there was little theoretical development on how individual-level training outcomes result in organizational-level outcomes (Tharenou, Saks, and Moore, 2007). Kraiger, McLinden, and Casper (2004) discussed the need for a theory that describes the link between training outcomes and organizational or business outcomes. In addition to, or maybe because of the missing theoretical foundation, the actual measurement of training effects on individual level and organization level outcomes asks for sophisticated econometric models that can deal with the various influences on the outcome level; these models should also, at the same time, quantify the effect of training. Evaluation studies on the results level are faced with diverse statistical difficulties, but microeconometric program evaluation literature can address these issues, when added to the existing managerial and psychological methods of training evaluation.

Microeconometric evaluation concepts

In the context of evaluation and evaluation research, two issues are encountered: (i) the fundamental evaluation problem in measuring the counterfactual outcome and (ii) the risk of an ensuing selection bias. In order to estimate the true treatment effect, some identifying assumptions are required to draw inference about the hypothetical population based on the observed population (Caliendo and Hujer, 2006). To solve the fundamental evaluation problem, (i.e. the problem that individuals may not be observed simultaneously with and without training participation), various methods are offered in the microeconometric evaluation literature (for an overview, see Caliendo and Hujer, 2006). When it comes to selection, i.e. distorted representation of a true population in a sample, the "gold standard" is considered to be random assignment to the treatment and comparison group (experimental data), as members of both groups would then be equal in expectation and identical on all observed and unobserved characteristics, on average (Lechner, 1998). However, what may be applied in the economics of education research, cannot be transferred to the concept of company training, since observational (quasi-experimental) data is usually present, which entails selection effects by the nature of the training. The comparison group approach by Leuven and Oosterbeek (2008) is closest to random selection in a firm environment. In order to ensure comparability in observational and un-observational variables of participants and non-participants, only employees who are enrolled in a training program but cancelled due to random reasons, are assigned to the comparison group of non-participants. Since this approach puts high requirements on the data, Görlitz (2011) and Fahr, Hinerasky, and Simons (2014) are the only researchers who were able to apply this concept to a company training context. Besides, Caliendo and Hujer (2006) and Lechner (1998) offered an extensive overview of methods to eliminate or minimize the selection bias. Table 1 presents an overview of different econometric and statistical methods and their data requirements.

The essence of the selection problem is a distorted representation of a true population in a sample as a consequence of the sampling rule (Heckman, 2001). Hence, some identifying assumptions have to be imposed, to draw inferences about the hypothetical population based on the observed population. Each statistical method has different data requirements and assumptions when it comes to handling selection, but two broad categories of estimators can be distinguished according to how selection bias is handled. The first category relies on the selection on observables assumption, where the estimator is based on some form of exogeneity (unconfoundedness assumption). In other words, the selection into treatment depends only on the observable characteristics, and unobservable characteristics are independent of the outcome. If this assumption cannot be justified, a second category of estimators explicitly allows for selection on unobservable characteristics (Instrumental Variable approach and Heckman Selection Estimator).

Table 1: Overview of common statistical methods for non-experimental data in the microeconometric evaluation literature

	Statistical Method	Data	Selection	Functional Form	Effect Heterogeneity
1	Before-after estimator (BA)	Cross-section	Not necessary	Parametric	Estimation among subpopulation
2	Linear Regression Approach	Cross-section	Accounts for Selection on observables	Parametric	Estimation among subpopulation
3	Matching estimator	Cross-section	Accounts for Selection on observables	Nonparametric	Possible
4	Difference-in-difference estimator (DID)	Longitudinal	Accounts for Selection on unobservables (only time-invariant)	(Semi) Parametric	Estimation among subpopulation
5	Instrumental variables (IV)	Cross-section	Allows for selection on unobservables	Parametric, Nonparametric possible[1]	Invalidates the application of IV
6	Heckman selection estimator	Cross-section	Allows for selection on unobservables	Nonparametric	Possible

Note: Based on Caliendo and Hujer (2006) and Lechner (1998). For more information on evaluation methods for non-experimental data, see Blundell and Costa Dias (2000). An application of the statistical methods for the estimation of treatment models with heterogeneous effects under observable and unobservable selection has recently been made available and easy to implement, e.g. by Stata's command of "ivtreatreg" (Cerulli, 2011). [1]For an explanation of the difference between a parametric and nonparametric IV approach, see Horowitz (2011) and Newey (2013) for nonparametric IV estimations.

The before-after estimator (1) uses the participant's observable outcome in the pre-training period to represent the participant's unobservable counterfactual outcome without training in the after-training period. This approach assumes no systematic differences in the observable variables between the treatment and control group and does not allow for a shift in outcomes over time. The basic idea of the cross-section estimator (2), on the other hand, is to compare the after-training outcomes of training participants with the outcomes of non-participants at the same time period. The matching estimator (3), a nonparametric approach, avoids functional form assumptions that are implicit in linear regression models, such as (1) and (2), but it requires information from participants and non-participants for the time before and after training participation. The matching approach identifies the training effect by assigning each participant a non-participant that is similar in all relevant pre-training characteristics. The differences in outcomes between the matched pairs can then be attributed to the training. To reduce the selection bias, the matching estimator needs a rich dataset in order to find a comparable non-participant for every combination of observable variables of the participants. Matching is entirely flexible with respect to observable variables, as it is irrelevant whether variables determining the outcome are also determining selection or how observable and unobservable variables relate to the potential outcome (Lechner, 2008). Hence, a distinction between observable variables determining outcome, observable variables determining selection, and relevant unobservable variables determining selection does not play a role.

Frequently, controlling for selection on observable variables is not sufficient, as the heterogeneity between participants in unobservable characteristics such as in their cognitive ability or motivation may still lead to a biased estimation of treatment effects. The same applies for unobserved variables that may not be collectable or present in the data. If these unobserved or unobservable variables drive the selection bias, there are three strategies: (i) selection models that try to explicitly model the selection process, (ii) Instrumental Variable (IV) methods that try to include a variable that affects the participation decision but not the outcome, and (iii) Difference-in-Difference (DID) methods that erase a time-invariant selection effect by differencing outcomes of participants and non-participants, before and after treatment occurred. For that reason, longitudinal data are necessary.

Evaluation in the context of company training

Every estimator makes some generally untestable assumptions in order to overcome the fundamental evaluation problem. Choosing the right identification strategy to obtain reliable estimates of the causal effect is very context specific (Lechner, 2008). For example, the flexibility and robustness of nonparametric identification and estimation strategies has the drawback of needing larger samples. When evaluating company training literature, the datasets used usually belong to one of the following two groups: studies that have a company dataset or studies that use representative (household) survey data. Although company data sets deal with rather small sample sizes, they offer rich individual and firm-specific data. Representative data, on the other hand, is of a mid to larger sample size, but is often less informative when it comes to details concerning the individual and firm environment. As evident from previous information, the transfer of training is different on an individual basis and depends heavily on the organizational environment. Such a context calls for a statistical method that still operates reliably in small datasets. An intermediate solution is seen in the econometric case study approach,[6] which involves a trade-off between internal and external validity (Jones, Kalmi, and Kauhanen, 2012) but enables the collection of a rich set of variables that broader studies omit. Once an econometric case study offers enough information to address the selection effects, such rich company data and firm insight provide a good opportunity to study treatment effects.

Due to the comparably small sample sizes,[7] the typical econometric strategy to evaluate company training is regression analysis, and in some cases, IV approaches. An application of matching approaches fails because of the high data requirements. While regression approaches can control for unobservable time-invariant influences in longitudinal data, the results can be biased when these unobservables, such as innate ability or cognitive and non-cognitive skills, are linked to the decision to participate, or at the same time influence the decision to participate, and the training outcome.

Expanding from selection on observables to selection on unobservables

In the training evaluation literature, unobservable characteristics such as innate ability or non-cognitive skills have been shown to influence training outcome on the level of individual performance (Hinerasky and Fahr, 2014b), on earnings (Eren and Ozbeklik, 2013), and industry wage differentials (Murphy and Topel, 1990). While unobservable characteristics bias estimates in the regression context, regression analysis nonetheless remains the preferred strategy in evaluating training with firm datasets, due to its innate specifics. Therefore, it would be interesting and important to quantify such estimate distortion. Since the selection bias arises because of unobservable characteristics, the calculation of these unobservable influences can only naturally be conjectured. Altonji, Elder, and Taber (2005, 2002)

[6] For a review of this approach, also called the „insider econometrics" methodology, see Ichniowski and Shaw (2003) or Jones and Kato (2011).
[7] The average sample size of company data sets in a meta-analysis by Tharenou, Saks, and Moore (2007) is N=271.

developed an estimation technique that may quantitatively assess the degree of omitted variables bias by proxying the amount of selection on the unobservables, based on the amount of selection on the observed explanatory variables. The selection bias can then be calculated as the ratio of selection on unobservables to the selection on observables. Using a bivariate probit model to assess "the sensitivity of the estimates to assumptions about the amount of selection on unobservables," bound estimates are provided based on assumptions about the ratio of selection on observables to unobservables (Altonji, Elder, and Taber, 2008, p. 345). The proposed theoretical foundation, however, cautions about the possibility to infer too much about selection on unobservables from the selection on observables if the number and the explanatory power of explanatory variables is small or "if they are unlikely to be representative of the full range of factors that determine the outcome" (Altonji, Elder, and Taber, 2005, p. 182). However, Altonji, Elder, and Taber's approach is an interesting and important way to gain insight in the selection biases and to quantitatively assess, at least to some extent, distortions from selection. When empirical training evaluation asks for a regression approach that does not allow for selection on unobservables to distinguish pure correlation from causal effects, Altonji, Elder and Taber's approach (Altonji, Elder, and Taber, 2008, 2005, 2002) can be highly valuable for classifying the obtained results. This approach has already been applied to the evaluation of youth training (Card et al., 2011), the effects of government sponsored training (Lechner, Miquel, and Wunsch, 2011), and the effect of outsourcing on workplace performance (Magnani, 2012). To our knowledge, however, it has yet to be applied in the context of company training.

Dynamic evaluation concepts

When it comes to the definition and measurement of the training variable, the training measure in representative household datasets is received from single-item answers ("Have you received training during the last 12 months?") or, in order to receive a more reliable measure, from aggregate measures ("How often have you participated in training during the last 12 months?"). In company datasets, the provision of training often starts and ends at the same point in time for all participants. In representative data, which covers various individuals in multiple companies, training participation may occur at different points in time; therefore, nonparametric regression methods may not be applied (Lechner, 1998). When an individual participates in subsequent treatments, e.g. decides to participate in training in period one and again in period two, the outcome in the second period will naturally be influenced by the participation decision in the first period. Therefore, the outcome in the second period is no longer only dependent on the explanatory variables (covariates X), but also dependent on the participation decision in period one and its related outcome. In the empirical training literature, aggregate measures are commonly used, but "most empirical work that contains dynamic selection problems ignores the intermediate outcomes and treats the sequence participation as being determined form the start" (Caliendo, 2006, p. 89). The empirical strategy should therefore consider dynamic evaluation concepts, three of which are briefly described in Table 2.

Table 2: Dynamic evaluation concepts

	Statistical Method	**Description**	**Functional Form Assumptions**
1	Duration Model Approach (Abbring and Berg, 2004, 2003)	Causal relation between two durations, outcome duration and treatment time	Nonparametric
2	Sequential Matching Estimators (Lechner and Miquel, 2010; Lechner, 2004)	Individuals can participate in subsequent treatments	Nonparametric
3	Matching with Time-Varying Treatment Indicators (Fredriksson and Johansson, 2008; Sianesi, 2004)	Takes timing of events into account as outcome of stochastic process	Nonparametric

Source: Based on Caliendo and Hujer (2006).

Each of the estimators also concern handling the selection bias and use nonparametric functional form assumptions for the outcome equation. An application in the evaluation literature assumes rich datasets, which allow the implementation of a matching strategy. Even more demanding than the estimation strategies presented in Table 1 is the requirement for trained and non-trained matches to not only share the same configuration in explanatory variables, but also to share the identical pre-training period lengths. However, the results of an application of dynamic evaluation concepts in the training evaluation for the unemployed (e.g. Fredriksson and Johansson, 2003, 2008) lead to some general suggestions for empirical work. First, information on the timing of the training event should not be discarded, but incorporated into the identification strategy. Second, and most importantly, information on multiple treatments should not be aggregated into a binary treatment indicator, as this leads to inference under much weaker assumptions (Abbring and Berg, 2004).

6 Return on Investment

Quantifying the results that stem from training participation has been shown to be complex. With the use of empirical methods, however, reliable estimates can be found to quantify the impact of training on a given outcome. Phillips (1996) added a sixth level to Kirkpatrick's taxonomy, namely the calculation of a training program's return on investment (ROI), by transforming individual and organizational level results into a monetary unit that is contrasted with the costs of intervention. Positive training results might turn out negative once their returns on investment are calculated. Such a calculation is, at first sight, valuable for human resource departments. However, it proves to be rather difficult. The average treatment effects of training can only be calculated for the period under observation and may be inaccurate due to a missing variable bias or selection bias. The various nature of training courses, such as career advancement, adaption of new skills, compensation of skill retention, etc. involve a tedious quantification of long-term results, which misguides the ROI calculation. The impact of training on the ROI is secondary for the firms' decision to provide training, since firms face a fast-paced industrial environment that urges them to train their employees in order to stay ahead in competition. In order for training to be successful, firms have to adapt them to their individual business requirements.

7 Conclusion

We review and combine the training evaluation literature conducted in different disciplines. Based on Kirkpatrick's goal-oriented evaluation taxonomy (Kirkpatrick, 1979), major insights from the different literature streams were summarized into each evaluation level, and the gaps between research conducted in separate fields were filled. Literature streams form psychology, education, economics, and statistics, enriched the content of Kirkpatrick's training evaluation taxonomy and identified opportunities for further development.

Learning and retention are positively influenced by three categories of training inputs: trainee characteristics, training design, and work environment characteristics (Saks, Salas, and Lewis, 2014). The extensive cognitive literature on the transfer of learning analyzes each training form with respect to its inherent learning process (see Daffron and North, 2011) and suggests ways to improve individual learning, which offers valuable insight to enhance the transfer of training. Each training form, i.e. action learning, e-learning, or team training, subsequently requires an individual approach to improve the transfer of learning within organizations (see Baldwin and Ford, 1988 and Holton and Baldwin, 2003). To facilitate learning transfer, the differences between training forms have to be considered. In order to facilitate transfer from e-learning, the objective must lie in the implementation and application of the learned content. This content is already incorporated into other forms of learning, such as action learning, which is also known as "learning by doing". Human Capital theory assumes that higher amounts of human capital inherently lead to higher productivity. This causality, however, has not always been confirmed in cases of company training (Görlitz, 2011; Hinerasky and Fahr, 2014b; Pischke, 2001). Insights into the learning process open this "black box" within Human Capital theory by explaining how an increase in productivity can result from learning and permanent transfer, and it provides important leverage in achieving the causal relation between learning and productivity through targeted training design.

While the immediate results of training are often measured positive after participation, a true challenge lies in the permanent transfer of learned material into the workplace. As the motivation to transfer, and transfer itself, are not only influenced by the individual factors but also complemented by environmental factors that constitute the organizational reality, each company has to design its individual transfer system. This insight is equally important for research scholars and economic decision-makers, whose (performance-oriented) interest lies in addressing low retention rates and the successful transfer of training, in order to keep a financial balance. Blume et al. (2010) surprisingly found only few individual features that are consistently strong in predicting transfer. However, "the reality [is] that there are no magic bullets for leveraging transfer“ (Blume et al., 2010, p. 1096). As work environment factors are just as strongly related to transfer as trainee characteristics, especially the transfer climate, supervisor support and peer support play an important role in improving transfer. Hence, companies have to analyze their internal structures to develop the most suitable path to facilitate and improve the transfer of training and ultimately enhance firm performance. When guiding organizations in this respect, they should expand their “happy sheets” when evaluating training programs by not only retrieving training satisfaction, but by asking more specific questions regarding training utility. Subjective assessments of the learning amount and training success should be replaced by objective assessments. Since the focus is on transfer, the application of training material can be accompanied by individual reminders or coaching and repeated evaluations. Immediate training effects that vanish shortly after training completion emphasize again the integration of the training content into regular work life and routines.

Encouraging the interaction of ideas, research methods, and results between economists and other scholars interested in the economic dimension of training, the role of econometric analysis in evaluating training programs is prominent. In order to quantitatively assess training effectiveness, applied micro-econometric methods are used to not only quantify the results but also to test underlying assumptions given from theoretical frameworks. As a result, we are not only able to highlight influencing factors when it comes to training success and to quantify their relative impact, but we are also able to assess skill decay after training participation. Future research lies in the integration of dynamic evaluation concepts and the adaption of selection on unobservables (Altonji, Elder, and Taber, 2008, 2005) into the standard regression analysis routine. This will help to quantify training results more precisely and to better understand the characteristics that lead individuals to participate in training, leading to the successful transfer of learning content. While empirical results offer new perspectives on the understanding and further development of the underlying theoretical models, educational and psychological insights into the learning process can be used to adapt training programs in order to support participants in their learning behaviors and to create a valuable base for the important transfer of knowledge into the workplace.

8 References

Abbring, J.H., Berg, G.J. van den, 2003. The Nonparametric Identification of Treatment Effects in Duration Models. Econometrica 71, 1491–1517.

Abbring, J.H., Berg, G.J. van den, 2004. Analyzing the effect of dynamically assigned treatments using duration models, binary treatment models, and panel data models. Empirical Economics 29, 5–20.

Alliger, G.M., Tannenbaum, S.I., Jr, W.B., Traver, H., Shotland, A., 1997. A meta-analysis of the relations among Training Criteria. Personnel Psychology 50, 341–358.

Altonji, J.G., Elder, T.E., Taber, C.R., 2002. An Evaluation of Instrumental Variable Strategies for Estimating the Effects of Catholic Schools.

Altonji, J.G., Elder, T.E., Taber, C.R., 2005. Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools. Journal of Political Economy 113, 151–184.

Altonji, J.G., Elder, T.E., Taber, C.R., 2008. Using Selection on Observed Variables to Assess Bias from Unobservables when Evaluating Swan-Ganz Catheterization. American Economic Review: Paper and Proceedings 98, 345–350.

Alvarez, K., Salas, E., Garofano, C.M., 2004. An Integrated Model of Training Evaluation and Effectiveness. Human Resource Development Review 3, 385–416.

Aragon, I.B., Valle, R.S., 2013. Does training managers pay off?. The International Journal of Human Resource Management 24, 1671–1684.

Arguinis, H., Kraiger, K., 2009. Benefits of Training and Development for Individual and Teams, Organizations, and Society. Annual Review of Psychology 60, 451–174. doi:10.1146/annurev.psych.60.110707.163505

Arthur, W., Bennett, W., Stanush, P., McNelly, T., 1998. Factors That Influence Skill Decay and Retention: A Quantitative Review and Analysis. Human Performance 11, 57.

Arthur, W.J., Bennett, W.J., Edens, P.S., Bell, S.T., 2003. Effectiveness of training in organizations: A meta-analysis of design and evaluation features. Journal of Applied Psychology 88, 234–245.

Baldwin, T., Ford, K., 1988. Transfer of Training: A Review and new Directions for future Research. Personnel Psychology 41, 63–105.

Becker, G., 1964. Human Capital. Columbia University Press, New York.

Becker, G., 1993. Human Capital. The University of Chicago Press.

Blume, B.D., Ford, J.K., Baldwin, T.T., Huang, J.L., 2010. Transfer of training: A meta-analytic review. Journal of Management 36, 1065–1105.

Blundell, R., Costa Dias, M., 2000. Evaluation Methods for Non-Experimental Data. Fiscal Studies 21, 427–468.

Brown, K.G., 2005. An Examination of the Structure and Nomological Network of Trainee Reactions: A Closer Look at 'Smile Sheets'. Journal of Applied Psychology 90, 991–1001.

Burke, L.A., Hutchins, H.M., 2007. Training transfer: An integrative literature review. Human Resource Development Review 6, 263–296.

Caliendo, M., 2006. Microeconometric Evaluation of Labour Market Policies, Lecture Notes in Economics and Mathematical Systems No. 568. Springer-Verlag, Berlin.

Caliendo, M., Hujer, R., 2006. The Microeconometric Estimation of Treatment Effects - An Overview. Journal of the German Statistical Society / Allgemeines Statistisches Archiv 90, 197–212.

Campbell, J.P., 1971. Personnel training and development. Annual Review of Psychology 22, 565–602.

Cannon-Bowers, J.A., Salas, E., Tannenbaum, S.I., Mathieu, J.E., 1995. Toward Theoretically Based Principles of Training Effectiveness: A Model and Initial Empirical Investigation. Military Psychology 7, 141–164. doi:10.1207/s15327876mp0703_1

Card, D., Ibarrarán, P., Regalia, F., Rosas-Shady, D., Soares, Y., 2011. The Labor Market Impacts of Youth Training in the Dominican Republic. Journal of Labor Economics 29, 267–300.

Cerulli, G., 2011. ivtreatreg: a new STATA routine for estimating binary treatment models with heterogeneous repsonse to treatment under observable and unobservable selection, in: 8thItalianStataUsersGroupmeeting.Venice,November1718.

Colquitt, J.A., LePine, J.A., Noe, R.A., 2000. Toward an Integrative Theory of Training Motivation: A Meta-Analytic Path Analysis of 20 Years of Research. Journal of Applied Psychology 85(5), 678–707.
Daffron, S., North, M., 2011. Successful Transfer of Learning (Professional Practices in Adult Education and Lifelong Learning). Krieger Publishing Company, Malabar, FL, USA.
DeGrip, A., Sauermann, J., 2013. The effect of training on Productivity: The transfer of on-the-job training from the perspective of economics. Educational Research Review 8, 28–36.
Dixon, N.M., 1990. The Relationship Between Trainee Responses on Participant Reaction Forms and Posttest Scores. Human Resource Development Quarterly 1, 129–137.
Eren, O., Ozbeklik, S., 2013. The Effect of Noncognitive Ability on the Earnings of Young Men: A Distributional Analysis with Measurement Error Correction. Labour Economics 24, 293–304.
Fahr, R., Hinerasky, C., Simons, S., 2014. Wage Returns of Company Training - Evidence from a Comparison Group Approach.
Ford, J.K., Kraiger, K., Merritt, S.M., 2010. An updated review of the multidimensionality of training outcomes: New directions for training evaluation research., in: Learning,training, developmentOrganizations, SIOPorganizationalFrontiersseries. Routledge/Taylor & Francis Group, pp. 135–165.
Fredriksson, P., Johansson, P., 2008. Dynamic Treatment Assignment: The Consequences for Evaluations Using Observational Data. Journal of Business & Economic Statistics 26, 435–445.
Gist, M.E., Stevens, C.K., Bavetta, A.G., 1991. Effects of Self-Efficacy and Post-training Intervention on the Acquisition and Maintenance of Complex Interpersonal Skills. Personnel Psychology 44, 837–861.
Goldstein, I.L., 1980. Training in work organizations. Annual Review of Psychology 31, 229–272.
Goodman, J.S., Wood, R.E., 2004. Feedback Specificity, Learning Opportunities, and Learning. Journal of Applied Psychology 89, 809–821.
Goodman, J.S., Wood, R.E., Chen, Z., 2011. Feedback specificity, information processing, and transfer of training. Organizational Behavior and Human Decision Processes 115, 253–267.
Görlitz, K., 2011. Continuous Training and Wages: An Empirical Analysis Using a Comparison-Group Approach. Economics of Education Review 30, 691–701.
Grossman, R., Salas, E., 2011. The transfer of training: what really matters. International Journal of Training & Development 15, 103–120.
Heckman, J.J., 2001. Micro Data, Heterogeneity, and the Evaluation of Public Policy: Nobel Lecture. Journal of Political Economy 109, 673–748.
Hinerasky, C., Fahr, R., 2014a. Learning Outcomes, Feedback, and the Performance Effects of a Training Program. Working Paper, University Paderborn, Germany.
Hinerasky, C., Fahr, R., 2014b. When the Early Bird Catches the Worm: The Impact of training in retail. Working Paper, University of Paderborn, Germany.
Holton, E.F., 2005. Holton's Evaluation Model: New Evidence and Construct Elaborations. Advances in Developing Human Resources 7, 37–54.
Holton, E.F., Baldwin, T., 2003. Improving Learning Transfer in Organizations. Jossey-Bass, San Francisco, USA.
Holton, I. Elwood F., 1996. The Flawed Four-Level Evaluation Model. Human Resource Development Quarterly 7, 5–21.
Holton, I. Elwood F., Bates, R.A., Ruona, W.E.A., 2000. Development of a Generalized Learning Transfer System Inventory. Human Resource Development Quarterly 11, 333–360.
Horowitz, J.L., 2011. Applied Nonparametric Instrumental Variables Estimation. Econometrica 79, 347–394.
Ichniowski, C., Shaw, K., 2003. Beyond Incentive Pay: Insiders' Estimates of the Value of Complementary Human Resource Management Practices. Journal of Economic Perspectives 17, 155–180.
Jones, D.C., Kalmi, P., Kauhanen, A., 2012. The effects of general and firm-specific training on wages and performance: evidence from banking. Oxford Economic Papers 64, 151–175. doi:doi:10.1093/oep/gpr013
Jones, D.C., Kato, T., 2011. The Impact of Teams on Output, Quality, and Downtime: An Empirical Analysis using Individual Panel Data. Industrial & Labor Relations Review 64, 215–240.

Kauffeld, S., Lorenzo, G., Weisweiler, S., 2012. Wann wird Weiterbildung nachhaltig?- Erfolg und Erfolgsfaktoren beim Lerntransfer. PERSONALquarterly 2, 10–15.
Kirkpatrick, D.L., 1979. Techniques for Evaluating Training Programs. Training & Development Journal 33, 78.
Kozlowski, S.W.J., Brown, K.G., Weissbein, D.A., Cannon-Bowers, J.A., Salas, E., 2000. A multilevel approach to training effectiveness: Enhancing horizontal and vertical transfer., in: Multileveltheory,Research, methodsOrganizationsFoundations,Extensions, new directions. Jossey-Bass, pp. 157–210.
Kraiger, K., Ford, J.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, 311–328.
Kraiger, K., McLinden, D., Casper, W.J., 2004. Collaborative Planning for Training Impact. Human Resource Management 43, 337–351.
Kraiger, K., Salas, E., Cannon-Bowers, J.A., 1995. Measuring knowledge organization as a method for assessing learning during training. Human Factors 37, 804–816.
Latham, G.P., 1988. Human Resource Training and Development. Annual Review of Ps 39, 545–582. doi:10.1146/annurev.ps.39.020188.002553
Leach, M.P., Liu, A.H., 2003. Investigating Interrelationships among Sales Training Evaluation Methods. The Journal of Personal Selling and Sales Management 4, 327–339.
Lechner, M., 1998. Qualifikation, Weiterbildung und Arbeitsmarkterfolg, in: Pfeiffer, F., Pohlheimer, W. (Eds.), ZEW-Wirtschaftsanalysen, Nomos Verlag.
Lechner, M., 2004. Sequential Matching Estimation of Dynamic Causal Models. IZA Discussion Paper No. 1042, Bonn.
Lechner, M., 2008. Some Thoughts on 'The Principles Underlying Evaluation Estimators with an Application to Matching' by James J. Heckman. Annales d'Economie et de Statistique 81–89.
Lechner, M., Miquel, R., 2010. Identification of the Effects of Dynamic Treatments by Sequential Conditional Independence Assumptions. Empirical Economics 39, 111–137.
Lechner, M., Miquel, R., Wunsch, C., 2011. Long-Run Effects of Public Sector Sponsored Training in West Germany. Journal of the European Economic Association 9, 742–748. doi:DOI: 10.1111/j.1542-4774.2011.01029.x
LeRouzie, V., Ouchi, F., Zhou, C., 1999. Measuring "What People Learned" versus "What People Say They Learned": Does the Difference Matter?, in: Paperpresented Annual MeetingAmericanEvaluationAssociation(Orlando,FL,November36,1999.
Leuven, E., Oosterbeek, H., 2008. An Alternative Approach to Estimate the Wage Returns to Private-Sector Training. Journal of Applied Econometrics 23, 423–434.
Liebermann, S., Hoffmann, S., 2008. The impact of practical relevance on training transfer: evidence from a service quality training program for German bank clerks. International Journal of Training & Development 12, 74–86.
Locke, E.A., Frederick, E., Lee, C., Bobko, P., 1984. Effect of self-efficacy, goals, and task strategies on task performance. Journal of Applied Psychology 69, 241–251.
Long, L.K., DuBois, C.Z., Faley, R.H., 2008. Online training: The value of capturing trainee reactions. Journal of Workplace Learning 20, 21–37.
Magnani, E., 2012. Vertical disintegration and training: evidence from a matched employer–employee survey. Journal of Productivity Analysis 38, 199–217. doi:DOI 10.1007/s11123-011-0256-9
Maier, N.R.F., 1973. Psychology in Industrial Organizations. Houghton Mifflin Company.
Mathieu, J.E., Tannenbaum, S.I., Salas, E., 1992. Influences of Individual and Situational Characteristics on Measures of Training Effectiveness. The Academy of Management Journal 35, 828–847.
Murphy, K.M., Topel, R.H., 1990. Efficiency Wages Reconsidered: Theory and Evidence., in: Advances theoryMeasurement unemployment. New York:, U Chicago, pp. 204–240.
Newey, W.K., 2013. Nonparametric Instrumental Variables Estimation. American Economic Review 103, 550–556.
Noe, R.A., Schmitt, N., 1986. The Influence of Trainee Attitudes on Training Effectiveness: Test of a Model. Personnel Psychology 39, Personnel Psychology.

Noe, R.A., Tews, M.J., McConnell Dachner, A., 2010. Learner Engagement: A New Perspective for Enhancing Our Understanding of Learner Motivation and Workplace Learning. Academy of Management Annals 4, 279–315.
Perkins, D.N., Salomon, G., 2012. Knowledge to Go: A Motivational and Dispositional View of Transfer. Educational Psychologist 47, 248–258.
Phillips, J.J., 1996. ROI: The Search for Best Practices. Training & Development.
Pischke, J.-S., 2001. Continuous Training in Germany. Journal of Population Economics 14, 523–548.
Quiñones, M.A., 1995. Pretraining context effects: Training assignment as feedback. Journal of Applied Psychology 80, 226–238.
Robertson, I., Downs, S., 1979. Learning and the Prediction of Performance: Development of trainability testing in the United Kingdom. Journal of Applied Psychology 64, 42–50. doi:10.1037/0021-9010.64.1.42
Rouiller, J.Z., Goldstein, I.L., 1993. The Relationship Between Organizational Transfer Climate and Positive Transfer of Training. Human Resource Development Quarterly 4, 377–390.
Rowold, J., 2007. Individual influences on knowledge acquisition in a call center training context in Germany. International Journal of Training & Development 11, 21–34.
Saks, A.M., Burke-Smalley, L.A., 2014. Is transfer of training related to firm performance?. International Journal of Training & Development 18, 104–115.
Saks, A.M., Salas, E., Lewis, P., 2014. The transfer of training. International Journal of Training & Development 2, 81–83.
Salas, E., Cannon-Bowers, J.A., 2001. The Science of Training: A Decade of Progress. Annual Review of Psychology 52, 471–499. doi:10.1146/annurev.psych.52.1.471
Salas, E., Tannenbaum, S.I., Kraiger, K., Smith-Jentsch, K.A., 2012. The science of training and development in organizations: What matters in practice. Psychological Science in the Public Interest 13, 74–101.
Segers, M., Gegenfurtner, A., 2013. Transfer of Training: New conceptuconceptual through integrated reseach perspectives. Educational Research Review 8, 1–4.
Sianesi, B., 2004. An Evaluation of the Swedish System of Active Labor Market Programs in the 1990s. Review of Economics and Statistics 86, 133–155.
Sitzmann, T., Brown, K.G., Casper, W.J., Ely, K., Zimmerman, R.D., 2008. A review and meta-analysis of the nomological network of trainee reactions. Journal of Applied Psychology 93, 280–295.
Tannenbaum, S., Cannon-Bowers, J.A., Mathieu, J.E., 1993. Factors That Influence Training Effectiveness: A Conceptual Model and Longitudinal Analysis. Naval Train. Sys. Cent.
Tannenbaum, S.I., 1997. Enhancing Continuous Learning: Diagnostic Findings from Multiple Companies. Human Resource Management 36, 437–452.
Tannenbaum, S.I., Mathieu, J.E., Salas, E., Cannon-Bowers, J.A., 1991. Meeting Trainees' Expectations: The Influence of Training Fulfillment on the Development of Commitment, Self-Efficacy, and Motivation. Journal of Applied Psychology 76(6), 759–769.
Tannenbaum, S.I., Yukl, G., 1992. Training and Development in Work Organizations. Annual Review of Psychology 43, 399–441.
Tharenou, P., Saks, A.M., Moore, C., 2007. A review and critique of research on training and organizational-level outcomes. Human Resource Management Review 17, 251–273.
Volet, S., 2013. Extending, broadening and rethinking existing research on transfer of training. Educational Research Review 8, 90–95.
Warr, P., Bunce, D., 1995. Trainee Characteristics and the Outcomes of Open Learning. Personnel Psychology 48, 347–375.
Weisweiler, S., Nikitopoulos, A., Netzel, J., Frey, D., 2013. Gaining insight to transfer of training through the lens of social psychology. Educational Research Review 8, 14–27.
Wexley, K.N., 1984. Personnel Training. Annual Review of Psychology 35, 519–551. doi:10.1146/annurev.ps.35.020184.002511