

## 60 Years Later - Testing Kirkpatrick's 4-Level Evaluation Model on an Extra-Occupational Master Program

Moritz Bauer, M.A.<sup>1</sup> & Michael Goehlich, Prof. Dr.<sup>1</sup>

### Abstract

---

This paper investigates the relationships between the evaluation levels in Kirkpatrick's (supposedly) hierarchically constructed 4-level model. Although this evaluation model is still regarded as the most popular and widespread, there are contradictory findings in the scientific discourse regarding the effects the individual evaluation levels have on each other. The present research thus contributes to sharpening the widely used evaluation model.

The study is based on a quantitative questionnaire assessing the four levels of the model – reaction, learning, (learning-)transfer and career-success. The questionnaire was completed by a total of 88 alumni of an extra-occupational part-time Master's program in the field of organizational and personnel development.

The data analysis using simple and multiple linear regressions shows the hierarchical structure of the model, as each evaluation level determines its subsequent level. This article thus confirms Kirkpatrick's modelling, first published 60 years ago, and contributes to a clearer understanding of the individual evaluation levels.

---

**Keywords:** evaluation, evaluation model, 4-level model, Kirkpatrick

### 1 Introduction

60 years ago, during the years 1959 and 1960, Donald Kirkpatrick published the principles of his evaluation model in a series of four articles (Kirkpatrick, 1959a, 1959b, 1960a, 1960b). Despite numerous more recent models (e.g. Day, Arthur, & Gettman, 2001; Hogan, Cepela, & Fentress, 2014; Holton III, 1996, 2005; Kauffeld, Bates, Holton III, & Müller, 2008; Kraiger, Ford, & Salas, 1993), Kirkpatrick's 4-level model is still considered to be the most popular and widely used one (Arthur et al., 2003; Gessler & Sebe-Opfermann, 2011; Salas & Cannon-Bowers, 2001; Praslova, 2010). Its high degree of popularity can be ascribed to its simplicity, plausibility and user-friendliness (Bates, 2004; Reio, Rocco, Smith, & Chang, 2017). Kirkpatrick did not intend to postulate a model in the scientific sense, but to formulate guidelines for carrying out meaningful evaluations (Kirkpatrick, 1979; Kraiger, 2003), as he stated that a distinct understanding could not be attributed to the term "evaluation": "I don't care whether it's a model or taxonomy as long as training professionals find it useful in evaluating training programs" (Kirkpatrick, 1996b, p. 55).

He proposed four levels for evaluation: *Reactions*, *Learning*, *Behavior* and *Results*. "By breaking it down in reaction, learning, behavior, and results, the training professional [...] can gradually progress from a simple subjective reaction sheet to a research design that measures tangible results" (Kirkpatrick, 1996a, p. 311). Each measure should be examined "at as many of the four levels as possible" (Kirkpatrick, 1996b, p. 57), even if this is more difficult and more expensive than examining individual levels (Kirkpatrick, 1979, 1987, 1996b, 1998; Kirkpatrick & Kirkpatrick, 2006). It should be noted that the degree of difficulty for the evaluation increases from level to level (Atria, Reimann, & Spiel, 2006; Gessler, 2005), which is why many evaluation studies only examine the levels *Reactions* and *Learning* (Gessler, 2005; Kirkpatrick & Kirkpatrick, 2006).

---

<sup>1</sup> Friedrich-Alexander-University Erlangen-Nuremberg, Chair of Education with a Focus on Organizational Education, Bismarckstraße 1a, D-91054 Erlangen, Germany

The extent to which the evaluation levels are related to each other or determine the subsequent evaluation levels has been widely discussed throughout the literature (e.g. Bates, 2004; Reio et al., 2017; Ritzmann, Hagemann, & Kluge, 2014; Salas and Cannon-Bowers, 2001) – however, without an unequivocal answer. Though Kirkpatrick's statements suggest a hierarchical conception of the model (albeit very vaguely), the scientific discourse – even 60 years after the publication of the model – contains contradictory findings on existing (non-)correlations (see Chapter 2.3). The present study therefore examines the question of whether and in what form the relationships between the evaluation levels postulated by Kirkpatrick exist. Specifically, it seeks answers to the following questions:

1. To what extent does the evaluation level *reactions* affect the evaluation levels *learning*, *transfer* and *career success*?
2. To what extent does the evaluation level *learning* affect the evaluation levels *transfer* and *career success*?
3. To what extent does the evaluation level *transfer* affect the evaluation level *career success*?

Accordingly, this study does not deal with a research gap in the classical sense (little or no empirical studies on a specific topic). Rather, as previously mentioned, its relevance lies in its contribution to a) refining and b) revising the content of the frequently used evaluation model. Kirkpatrick's fourth level results, which means consequences on an organizational level (see Chapter 2.1.4), can only be examined if the participants are part of the same organization. However, as participant groups are often composed independently of their organizational affiliation, we replace results with career success and thus enable the investigation of the fourth evaluation level independently of the participants' organizational affiliation (e.g. to assess the impact of professional development – see Huber, 2011). Furthermore, all four evaluation levels now measure changes at a personnel level.

## 2 Kirkpatrick's 4-Level Evaluation Model

Between November 1959 and February 1960, Kirkpatrick (1959a, 1959b, 1960a, 1960b) published the fundamentals of his evaluation model developed in 1954 for the first time. In it he proposes carrying out evaluations on four levels: *reaction*, *learning*, *behavior* and *results*.

The *reactions* level is intended to examine the reactions of the participants to the training (Kirkpatrick, 1959a, 1970, 1996a, 1996b, 1998; Kirkpatrick & Kirkpatrick, 2006) and can also be described as customer satisfaction (Kirkpatrick & Kirkpatrick 2006). This level is particularly important because managers often base training decisions on participant's comments (Kirkpatrick 1996b, 1998).

The understanding of the second level *learning* has changed throughout the years. In his first publication Kirkpatrick (1959b) defines learning "in a rather limited way" (Kirkpatrick 1979, p. 82, 1987, p. 8) as "what principles, facts and techniques were understood and absorbed by the conferees?" (Kirkpatrick 1979, p. 82, 1987, p. 8). Later, however, learning is understood as "the extent to which participants change attitudes, improve knowledge, and/or increase skill as a result of attending the program" (Kirkpatrick & Kirkpatrick, 2006, p. 22). It must be emphasized that the above definition includes "statements on the reception, processing and management of learning content and principles by participants" (Tirre, 2012, p. 21), but not the application of what has been learned, i.e. a change in behavior (Kirkpatrick, 1996b, 1998; Kirkpatrick & Kirkpatrick, 2006; Tirre, 2012).

The *behavior* level describes the extent to which a participant's work behavior has changed due to the training (Kirkpatrick, 1960a, 1996b, 1998; Kirkpatrick & Kirkpatrick, 2006). A total of four conditions are necessary to induce this behavioral change, whereby only the first two can be influenced by training: (1) the desire for change, (2) the knowledge of what to do and how to do it, (3) the appropriate working climate and (4) intrinsic (e.g. feelings of satisfaction, pride or success) as well as extrinsic (e.g. praise from superiors, salary increases) rewards for the change (Kirkpatrick & Kirkpatrick, 2006).

*Results* are the consequences of attending a training course at the organizational level (Kirkpatrick 1960b), and "the reason for having some training programs" (Kirkpatrick & Kirkpatrick, 2006, p. 25). Therefore, they must be consistent with the respective training objectives (Kirkpatrick, 1960b; Kirkpatrick & Kirkpatrick, 2006). Results can include, for example, increased productivity, improved quality, reduced costs, reduced accident rates and/or severity, improved sales figures, lower turnover and higher profits (Kirkpatrick, 1996b, 1998; Kirkpatrick & Kirkpatrick, 2006). Contrary to earlier publications (among others Kirkpatrick, 1960b, 1996a, 1998), Kirkpatrick and Kirkpatrick (2006) now state that not all of these results must be measured in concrete numbers. Therefore, it does not necessarily require quantitative data collection – nevertheless that should be the aim.

Accordingly, collecting data for the results-level is the most challenging part of the training evaluation (Gessler, 2005; Kirkpatrick & Kirkpatrick, 2006), however it is also the most important for those responsible in the company of the participants (Kirkpatrick & Kirkpatrick, 2006).

## 2.1 Operationalization in the present study

Alliger et al. (1989; 1997) investigated Kirkpatrick's 4-level model in two meta-analyses. In their first analysis, based on 12 publications and 26 correlations, they focused primarily on the relationships between the levels. The second analysis, however, is characterized by the fact that the respective levels could be specified in terms of their content and clearly distinguished from each other. On the basis of a total of 34 studies and 115 correlations, Alliger et al. (1997) propose an extension to Kirkpatrick's model.

This deeper understanding of the levels provides the basis for the operationalization in the present study, although Alliger et al. (1997) classifications are not followed in full. From our point of view, further refining of the model in certain places is necessary in order to be able to validly measure the respective levels.

### 2.1.1 Level 1 – Reactions

Alliger et al. (1997) divide the *reactions*-level into affective reactions and utility judgements. The sublevel affective reactions measures the satisfaction of the participants with the intervention (e.g. "I liked the training"). Utility judgements refer to assessments of usefulness (e.g. "Was the training of practical value?"). Alliger et al. (1997) point out that this differentiation was empirically based and independent from other authors, however, a similar differentiation (enjoyment of training - usefulness of training) can be found in Warr and Bunce (1995).

A third category, difficulty of training, highlighted by Warr and Bunce (1995), is not taken into account in the meta-analysis by Alliger et al. (1997), as "training difficulty seems to be rarely asked of trainees" (Alliger et al., 1997, p. 344). To fill this gap, we include this category in the evaluation model. Therefore, the reaction level in our study is examined by assessing *pleasure*, *usefulness* and *difficulty*.

### 2.1.2 Level 2 – Learning

Alliger et al. (1997) distinguish between three subcategories of *learning*: immediate post-training knowledge, which is knowledge measured directly after the intervention; knowledge retention, which describes knowledge acquired at a later point in time; and behavioral/skill demonstration, which includes all changes in behavior that occur due to the training.

Of interest here is that only the learning results are at the center of the construct description. Contrary to this, the present study follows Göhlich and Zirfas (2007, p. 17), where learning is understood as "an experiential reflection of the acquisition of specific knowledge and skills that affect the learner". Dialogicity is modally essential here, since learning essentially takes place in confrontation with the other (content of learning) and the others (co-learners, docents/trainers, etc.). Although learning in the current study cannot be captured in this complexity, the aspects *learning through case studies*, *learning through a practice-based project* and *learning through/with peers* at least attempt to take skill-learning as well as the dialogicity of learning into account. In addition, *general learning* and the *learning of subject-specific competencies* as well as *personal competencies* are examined.

### 2.1.3 Level 3 – Transfer

The third evaluation level, described as *behavior* by Kirkpatrick (1960a), is labeled by Alliger et al. (1997) as the *transfer*-level. They understand transfer as a more general change in behavior that occurs some time after the intervention and at the workplace. However, it remains unclear what exactly the cause for the change in behavior is. Different definitions exist in the scientific discussion of this term in the learning context (also synonymous with learning transfer). Bartz (2019, p. 45) understands transfer of learning as "an organization (*sic.*) member's ability to take knowledge and skills acquired in one context [...] and effectively transfer and apply the learning to new tasks or unfamiliar situations". In Baldwin, Ford and Blume (2009) one can also find a time-related component. However, they confine to learned behavior. They explain that "for transfer to have occurred, learned behavior must be generalized to the job context and maintained over a period of time on the job" (Baldwin et al. 2009, p. 41).

According to our understanding of learning (as defined above), it is assumed that both knowledge and skills are learned. In this respect, both aspects must be included in a definition of transfer and in the corresponding methodological instruments. Modally, the essential difference between learning and transfer is the time component: successful transfer occurs when the intended effects occur over a longer time period.

Therefore, stable changes over a longer period of time are relevant for the present understanding of the term transfer. Transfer is thus understood as the application and generalization (potentially even to changes in behavior shown over a longer period of time at the workplace) of the conveyed knowledge and skills. Therefore, we operationalize transfer via *transfer implementation of the alumni concerning their study content on the topics of organizational and personnel development, general transfer implementation* as well as *intrinsic and extrinsic transfer motivation of the alumni*.

#### 2.1.4 Level 4 – Career-Success

Both in Kirkpatrick's 4-level model (1960b) and in the extension of that model by Alliger et al. (1997), the fourth evaluation level is described as *results*. In their meta-analysis, Alliger et al. (1997) found only three studies that examine correlations with the fourth level. Therefore, they do not specify level 4 but rather follow Kirkpatrick's (1960b, 1996b, 1998; Kirkpatrick & Kirkpatrick, 2006) understanding of results as consequences on an organizational level (such as increased productivity, improved quality, reduced costs, higher profits).

Measuring success at the organizational level requires a) a sufficiently large number of participants from the same organization attending the training and b) that the specific organization to provide the relevant data (e.g. productivity rates, balance sheets, etc.). However, in our study no company was represented by more than one participant (see Chapter 3.1). Since the alumni all come from different organizations, no changes at the organizational level caused by the Master's program are to be expected. It is assumed, however, that graduating from the study program will produce "individual" results. In this respect, the personal *career success* of the alumni was examined.

Although career-success is a subject of many studies (Arthur, Khapova, & Wilderom, 2005; Aryee, Chay, & Tan, 1994; Blankenship, 1973; Boudreau, Boswell, & Judge, 2001; Gunz & Heslin, 2005; Hughes, 1937; Seibert & Kraimer, 2001), the definition of the term is vague, especially in the German language (Dette, Abele, & Renner, 2004). Difficulties are caused, among other things, by translation problems from English into German. Career means both the continuous promotion of a person – in German "Karriere" – and the descriptive term for the professional life of a person (e.g. change of company or job, periods of unemployment etc.) – in German "Laufbahn" (Abele, Spurk, & Volmer, 2011; Dette et al., 2004; Super, 1980). For example, Ng and Feldman (2014, p. 170) describe career as "the unfolding sequence of a person's work experiences over time and across multiple jobs, organizations, and occupations". Likewise, one can attribute a descriptive as well as an evaluative character to the English word success (German: "Erfolg"). It should be noted that success is a subjective construction (Abele et al., 2011). In this sense, career-success is defined as "the real or perceived achievements individuals have accumulated as a result of their work experiences" (Judge, Higgins, Thoresen, & Barrick, 1999, p. 622).

While earlier studies have focused on career-success primarily in the form of neutral indicators such as salary, hierarchy level or number of employees, the authors of more recent publications largely agree that career-success should be measured in terms of *objective* and *subjective success* (Barthauer, Estel, Dubbel, Kauffeld, & Spurk, 2016; Dette et al., 2004; Judge & Kammeyer-Mueller, 2007; Weber, 2013). We follow this differentiation.

Objective career-success is directly observable, measurable and verifiable by a third party (Abele et al., 2011; Barthauer et al., 2016; Heslin, 2005; Hughes, 1937). Typical criteria evaluated here are income, hierarchical position or the number of promotions (Abele et al., 2011; Barthauer et al., 2016; Heslin, 2005; Ng, Eby, Sorensen, & Feldman, 2005; Weber, 2013).

Subjective career-success, on the other hand, describes the individual assessments or evaluations of one's career progress so far and is the result of a comparison process (Abele, Hagmaier, & Spurk, 2015; Abele et al., 2011; Barthauer et al., 2016; Heslin, 2005; Hughes, 1937; Ng & Feldman, 2014). Such comparative judgements can either be made on a self-referential basis – i.e. in comparison with self-imposed norms and expectations, past performance and future goals – or on a cross-comparative basis – i.e. in comparison with other norms, reference persons or groups and social standards (Abele et al., 2011; Ng & Feldman, 2014).

## 2.2 Intercorrelations among the four Levels and Hypotheses

Not only have the definitions of the individual levels been discussed since the first publication of the 4-level model, but also the question of how they relate to each other. Kirkpatrick's statements (e.g. Kirkpatrick & Kirkpatrick, 2006) suggest that the model is designed as a hierarchical one and that the evaluation levels are interdependent (see e.g. also Bates, 2004, p. 342; Gessler & Sebe-Opfermann, 2011, p. 272), however these statements are rather vague. He assumes that "if training is going to be effective, it is important that trainees react favorably to it" (Kirkpatrick & Kirkpatrick, 2006, p. 27) and "without learning, no change in behavior will occur" (Kirkpatrick & Kirkpatrick, 2006, p. 50). Here "each level is important and has an impact on the next level" (Kirkpatrick & Kirkpatrick, 2006, p. 21). The empirical examination of these relationships does also not provide clear findings.

However, it is generally agreed upon that there is no connection between the first evaluation level (*reactions*) and the others (Alliger & Janak, 1989; Alliger et al., 1997; Arthur, Bennett, Edens & Bell, 2003; Colquitt, LePine, & Noe, 2000; Gessler & Sebe-Opfermann, 2011; Ruona, Leimbach, Holton III & Bates, 2002). Merely Saks and Burke (2012) demonstrate a medium correlation between the first and second level.

Some authors (Alliger & Janak, 1989; Alliger et al., 1997; Colquitt et al., 2000; Gessler & Sebe-Opfermann, 2011; Saks & Burke, 2012) identify correlations between the second evaluation level *learning* and the third evaluation level *transfer*. In Campion and Campion (1987), however, the learning-level cannot predict a positive change in the transfer-level.

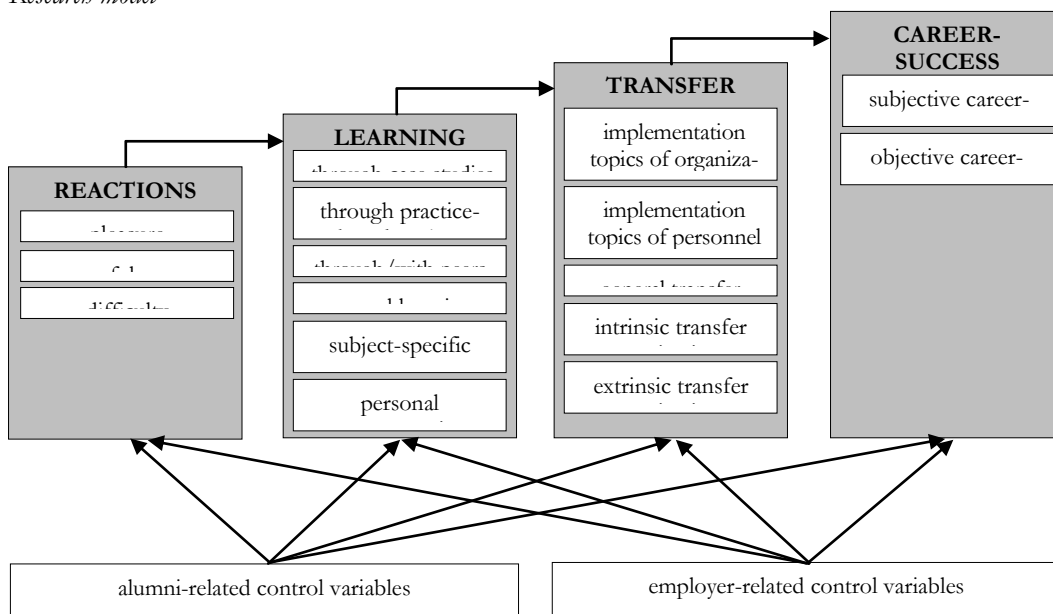
Since the fourth evaluation level is seldomly examined, there are only a few findings on the link between this level and the others. Alliger and Janak (1989) and Alliger et al. (1997) as well as Saks and Burke (2012) are able to demonstrate a connection between the levels *learning* and *results*. However, Alliger et al. (1997, p. 350) point out that "there were, however, only two studies reporting such correlations". Alliger and Janak (1989) as well as Alliger et al. (1997) provide identical findings for the relationship between the third and fourth evaluation levels, whereby in both publications only three studies report a correlation between the two levels.

In summary, it must be emphasized that the authors of the above-mentioned literature make use of different definitions and operationalizations, which makes generalization of the findings, in addition to the partly contradictory findings, impossible. Aware of its questionability and searching to get more clarity about its correctness, we base our research model (see Figure 1) and the following hypotheses on Kirkpatrick's hierarchical conception:

- H1: The evaluation level reactions has a positive impact on the evaluation levels learning, transfer and career-success.
- H2: The evaluation level learning has a positive impact on the evaluation levels transfer and career-success.
- H3: The evaluation level transfer has a positive impact on the evaluation level career-success.

Figure 1

Research model



### 3 Methods

#### 3.1 Research Context and Sample

In our evaluation study graduates of an extra-occupational part-time study-program at a university in southern Germany were interviewed. The interdisciplinary degree program was established by a university's institute of education in cooperation with a non-profit institute for vocational training and research. In line with the focus on organizational and personnel development, the degree program is aimed at university graduates from various disciplines whose fields of activity are already in the field of organizational and/or personnel development or who want to take on such an activity. To take part in this program, participants must have at least one year of professional experience after their first degree.

Of the 142 alumni that had completed their studies at the time of the survey, 88 took part in the survey. 67.00 % of the participants were female, 31.80 % were male (one person did not provide any information). They were on average 39.76 years old ( $SD = 8.60$ ) and one half of them had children.

Prior to the part-time master's program, 59 participants already held a diploma, four a magister, 17 a bachelor's degree and two a master's degree. One participant indicated the state examination as his or her highest university degree, while another named the church examination. Three interviewees had already obtained a PhD.

The graduates achieved a final grade of 1.73 ( $SD = .37$ ) on average, 1.0 being named as the best and 3.0 as the worst. Female alumni achieved significantly better marks with  $t(1.67) = 2.80$  ( $p < .01$ ) than the male participants.

The other sociodemographic data collected had no significant influence on the participants response behavior.

#### 3.2 Measures

The data was collected using a quantitative online questionnaire. The alumni were asked to (critically) evaluate the study program and thus support its further development.

According to the evaluation model, the research instrument consisted of four scales and background variables.

##### 3.2.1 Reactions-Scale

*Reactions* were measured using a self-constructed scale consisting of three subscales. The individual items of the subscales were developed based on item formulations of already existing measures (Gläßer et al., 2002; Grohmann & Kauffeld, 2013; Rindermann, 2009; Warr, Allan, & Birdi, 1999; Warr & Bunce, 1995), however, they were assembled in this form for the first time and adapted to the study context.

The subscales *pleasure* and *usefulness* (see 2.2.1. affective reactions and utility judgements) were assessed using a 5-point Likert-scale (disagree completely - agree completely), the subscale *difficulty*, on the other hand, used a two-dimensional, 7-point response scale (much too easy - exactly right - much too difficult).

A principal component analysis (PCA) was conducted on the 11 items with varimax (orthogonal) rotation. The Kaiser-Meyer-Olkin criterion verified the sampling adequacy for the analysis ( $KMO = .85$ ) which is well above the acceptable limit of .5 (Field, 2018). Bartlett-test of sphericity  $\chi^2(55) = 417.57$ ,  $p < .001$  reports that the correlations between the items are sufficiently large for PCA. An analysis of the eigenvalues shows that three components have values above the Kaiser criterion of 1 and in combination explain 69.16% of the variance. Table 1 shows the factor loadings after rotation.

Table 1 Summary of principal component analysis results for the reactions-scale

| Item          | rotated factor loadings |            |            |
|---------------|-------------------------|------------|------------|
|               | pleasure                | usefulness | difficulty |
| Ve1           | .78                     |            |            |
| Ve3           | .71                     |            |            |
| Ve4           | .71                     |            |            |
| Nue1          |                         | .75        |            |
| Nue2          |                         | .73        |            |
| Nue3          |                         | .70        |            |
| Nue5          |                         | .83        |            |
| SW1           |                         |            | .65        |
| SW2           |                         |            | .74        |
| SW3           |                         |            | .78        |
| SW4           |                         |            | .82        |
| Eigenvalues   | 4.53                    | 2.07       | 1.01       |
| % of variance | 26.80                   | 21.32      | 21.04      |
| $\alpha$      | .78                     | .86        | .76        |

Notes: lower factor loadings than those shown are suppressed

### 3.2.2 Learning-Scale

*Learning* is operationalized using six subscales. In addition to three self-constructed subscales – *learning through case studies*, *learning through the practice-based project*, and *learning through/with peers*, which follow the learning modal meaning of experience and dialogicity in knowledge and skill learning – a subscale for *general learning*, which goes beyond the contents of the first three subscales, was adapted. This subscale is taken from the Students' Evaluation of Educational Quality Questionnaire (Marsh 1982), translated into German via a forward-back-translation process and adapted to the evaluation objective. In addition, *subject-specific competencies* and *personal competencies* are assessed. These two subscales originate from the Berlin evaluation instrument for self-assessed student competencies (Braun, Gusy, Leidner, & Hannover, 2008). Only their wording was changed to fit the evaluation objective.

All items had to be answered using a 5-level Likert-scale (disagree completely- agree completely).

In contrast to the six theoretically formed subscales, the PCA with varimax rotation suggested a 5-factor structure (general learning could not be validated). The KMO criterion confirms sampling adequacy with  $KMO = .78$ . The Bartlett-test with  $\chi^2(210) = 1011.66$ ,  $p < .001$  also confirms sufficiently large correlations between the 21 items. As mentioned above, however, the analysis of the eigenvalues shows that only five, instead of the assumed six, factors can be extracted. These five components explain 68.45% of the variance. Table 2 reports the factor loadings after rotation.

Table 2 Summary of principal component analysis results for the learning-scale

| Item          | rotated factor loadings          |                                       |                                 |                                  |                           |
|---------------|----------------------------------|---------------------------------------|---------------------------------|----------------------------------|---------------------------|
|               | learning through<br>case studies | learning through<br>practical project | learning through/<br>with peers | subject-specific<br>competencies | personnel<br>competencies |
| PrL1          | .79                              |                                       |                                 |                                  |                           |
| PrL2          | .85                              |                                       |                                 |                                  |                           |
| PrL4          | .74                              |                                       |                                 |                                  |                           |
| PrL5          |                                  | .86                                   |                                 |                                  |                           |
| PrL6          |                                  | .88                                   |                                 |                                  |                           |
| PrL7          |                                  | .81                                   |                                 |                                  |                           |
| PrL8          |                                  | .84                                   |                                 |                                  |                           |
| PrL10         |                                  |                                       | .73                             |                                  |                           |
| PrL13         |                                  |                                       | .90                             |                                  |                           |
| PrL14         |                                  |                                       | .70                             |                                  |                           |
| Ler4          |                                  |                                       |                                 | .72                              |                           |
| Ler5          |                                  |                                       |                                 | .72                              |                           |
| FaKo1         |                                  |                                       |                                 | .78                              |                           |
| FaKo2         |                                  |                                       |                                 | .67                              |                           |
| FaKo3         |                                  |                                       |                                 | .76                              |                           |
| FaKo4         |                                  |                                       |                                 | .73                              |                           |
| PeKo1a        |                                  |                                       |                                 |                                  | .80                       |
| PeKo1b        |                                  |                                       |                                 |                                  | .80                       |
| PeKo2         |                                  |                                       |                                 |                                  | .66                       |
| PeKo3         |                                  |                                       |                                 |                                  | .63                       |
| PeKo4         |                                  |                                       |                                 |                                  | .65                       |
| Eigenvalues   | 6.57                             | 2.50                                  | 2.24                            | 1.95                             | 1.12                      |
| % of variance | 17.67                            | 15.75                                 | 13.71                           | 11.48                            | 9.83                      |
| $\alpha$      | .83                              | .90                                   | .74                             | .86                              | .80                       |

Notes: lower factor loadings than those shown are suppressed

### 3.2.3 Transfer-Scale

Five subscales were developed to measure *transfer*. Two subscales assess transfer implementation of the alumni concerning their study content on the topics of organizational and personnel development. The wording of the items was developed based on existing instruments (Gegenfurtner, 2013; Grohmann & Kauffeld, 2013; Kauffeld, Bates, Holton III, & Müller, 2008), but was strongly adapted to the object of evaluation and put together in this form for the first time.

For general transfer implementation, a subscale was designed based on items by Xiao (1996) and Machin and Fogarty (2004). The English-language items were translated by forward-back-translation into German and adapted to the object of evaluation.

In addition, two subscales examine the intrinsic and extrinsic transfer motivation of the alumni. This is done using a questionnaire from Gegenfurtner (2013).

All items were answered using a 5-point Likert-scale (disagree completely - agree completely).

The factor-analytical assessment first shows that with  $KMO = .74$  the sample adequacy is given and with  $\chi^2(153) = 896.03$ ,  $p < .001$  all items correlate sufficiently strong with each other. The analysis of the eigenvalues using PCA and varimax rotation, however, yields a 4-factor solution - the component of general transfer cannot be proven and must therefore be excluded from further investigations - which explains 67.23% of the variance. Table 3 shows the factor loadings after rotation.



Table 3 Summary of principal component analysis results for the transfer-scale

| Item          | rotated factor loadings   |  |                                  |                                  |
|---------------|---|--|----------------------------------|----------------------------------|
|               | transfer implement-<br>tation topics of orga-<br>nizational development | transfer implement-<br>tation topics of per-<br>sonnel development | intrinsic transfer<br>motivation | extrinsic transfer<br>motivation |
| TR1a          | .79   |  |                                  |                                  |
| TR2a          | .74   |  |                                  |                                  |
| TR3a          | .69   |  |                                  |                                  |
| TR5a          | .85   |  |                                  |                                  |
| TR6a          | .82   |  |                                  |                                  |
| TR1b          |   | .74  |                                  |                                  |
| TR2b          |   | .64  |                                  |                                  |
| TR3b          |   | .82  |                                  |                                  |
| TR5b          |   | .87  |                                  |                                  |
| TR6b          |   | .81  |                                  |                                  |
| iTm1a         |   |  | .57                              |                                  |
| iTm1b         |   |  | .65                              |                                  |
| iTm2a         |   |  | .82                              |                                  |
| iTm2b         |   |  | .84                              |                                  |
| eTm1a         |   |  |                                  | .56                              |
| eTm1b         |   |  |                                  | .82                              |
| eTm2a         |   |  |                                  | .74                              |
| eTm2b         |   |  |                                  | .70                              |
| Eigenvalues   | 6.41  | 2.37   | 2.06                             | 1.26                             |
| % of variance | 20.20   | 19.73  | 14.45                            | 12.85                            |
| $\alpha$      | .88   | .86  | .86                              | .74                              |

Notes: lower factor loadings than those shown are suppressed

### 3.2.4 Career-Success-Scale

*Career-success* of the interviewed alumni was surveyed using a questionnaire from Abele and various co-authors (Abele & Hagmaier, 2011; Abele et al., 2015, 2011; Spurk, Abele, & Volmer, 2011). *Career-success* is divided into *subjective* and *objective career-success*.

The subscale of *subjective career-success* is based on the Career Satisfaction Scale (Greenhaus, Parasuraman, & Wormley, 1990), but was extended by Abele et al. (Abele & Hagmaier, 2011; Abele, Hagmaier, & Spurk, 2015; Spurk et al., 2011) to include further items. It describes the subjective evaluation of career development (as career satisfaction and as comparative judgement). The subscale of *objective career-success*, on the other hand, measures a person's career performance in terms of monthly gross income (1 item) and hierarchical position (3 items). Both subscales could be replicated in several studies (Abele & Spurk, 2009; Abele, Spurk, & Volmer, 2011; Spurk & Abele, 2014; Spurk et al., 2011) and their statistical fit could be verified.

The items on the subscale of *subjective career-success* were collected using a 5-step Likert-scale (disagree completely - agree completely or less successful - more successful), whereas *objective career-success* was determined by a classification into income classes and closed yes-no questions.

Due to the answer formats, only the subscale of *subjective career-success* could be tested using reliability and factor analysis. PCA without rotation (if only one factor is to be extracted, no factor rotation is necessary) confirmed the 1-factor solution. The KMO criterion indicates sampling adequacy with  $KMO = .75$ . The Bartlett-test with  $\chi^2(21) = 256.05$ ,  $p < .001$  verifies as well sufficiently large correlations between the 7 items. The single component explains 48.81% of the variance. Table 4 reports the factor loadings.

Table 4 Summary of principal component analysis results for the subjective career-success-scale

| Item          | factor loadings           |
|---------------|---------------------------|
|               | subjective career-success |
| sKarr1        | .88                       |
| sKarr2        | .86                       |
| sKarr3        | .64                       |
| sKarr4        | .55                       |
| sKarr5        | .71                       |
| sKarr6        | .62                       |
| sKarr7        | .55                       |
| Eigenvalue    | 3.42                      |
| % of variance | 48.81                     |
| $\alpha$      | .82                       |

### 3.2.5 Background Variables

A total of 21 background variables were collected in the questionnaire. In addition to socio-demographic data (12 items), such as age or gender, the significance of the study program (4 items), the employer's support of the alumni during their studies (3 items) and a possible change of job or company (2 items) were assessed.

## 4 Results

As Table 5 shows, the *reaction*-, *learning*- and *transfer*-levels correlate weak to medium. Furthermore, the *transfer*-level shows a weak correlation with the *subjective career-success*. This in turn is related to the *objective career-success*. In the analysis, the evaluation level *career-success* is divided into the two separate scales *objective* and *subjective success*, which are considered separately, since the different scale levels do not allow a combination into a common scale.

Table 5 Correlation coefficients of the evaluation levels

|                             | <i>M</i> | <i>SD</i> | 1      | 2      | 3     | 4     |
|-----------------------------|----------|-----------|--------|--------|-------|-------|
| 1 Reactions                 | 4.34     | .54       | -      |        |       |       |
| 2 Learning                  | 3.90     | .45       | .60*** | -      |       |       |
| 3 Transfer                  | 3.49     | .60       | .41*** | .57*** | -     |       |
| 4 Objective Career-Success  | 3.39     | 1.10      | -.05   | -.21   | .12   | -     |
| 5 Subjective Career-Success | 3.52     | .81       | .06    | .13    | .30** | .39** |

Notes: \*\* $p < .01$ , \*\*\* $p < .001$

However, because correlations can only prove the existence and strength of a relationship between at least two variables, but are unable to model its nature (Rudolf & Müller, 2012), simple and multiple linear regression analyses were calculated to assess the relationships within the research model.

The analysis of the relationship between *reactions* and *learning* shows that the *reactions*-level has a positive influence on the *learning*-level,  $\beta = .60$  ( $p < .001$ ), and explains 35 % of the variance. Similarly, one can find a positive relationship between *reactions* and *transfer*,  $\beta = .41$  ( $p < .001$ ), whereby the predictor variable explains 16 % of the variance. Furthermore, the analysis of the relationship between the *learning*-level and the *transfer*-level also shows a positive influence of the predictor variable on the criterion variable with  $\beta = .57$  ( $p < .001$ ) while explaining 32 % of the variance.

Since a direct as well as an indirect link between the *reaction*-level and the *transfer*-level could be found, a further investigation into whether this relationship is mediated by the *learning*-level as a mediator variable must be conducted. Such a mediation occurs when the relationship to be mediated no longer emerges as significant when the mediator is controlled for (Baron & Kenny, 1986).

The multiple regression analysis shows (see Table 6) that by adding the evaluation level *learning*, the relationship between the *reaction*-level and the *transfer*-level is no longer significant. The calculation of the Sobel test shows a significant indirect effect ( $t_{\text{Sobel}} = 4.60, p < .001$ ) at the *learning*-level and thus proves the total mediation of the relationship between the *reaction*- and *transfer*-levels.

Table 6 Multiple regressions to test the learning level as a mediator variable

|           | B   | SE  | $\beta$ |
|-----------|-----|-----|---------|
| Step 1    |     |     |         |
| Reactions | .45 | .11 | .41***  |
| Step 2    |     |     |         |
| Reactions | .11 | .12 | .10     |
| Learning  | .69 | .15 | .51***  |

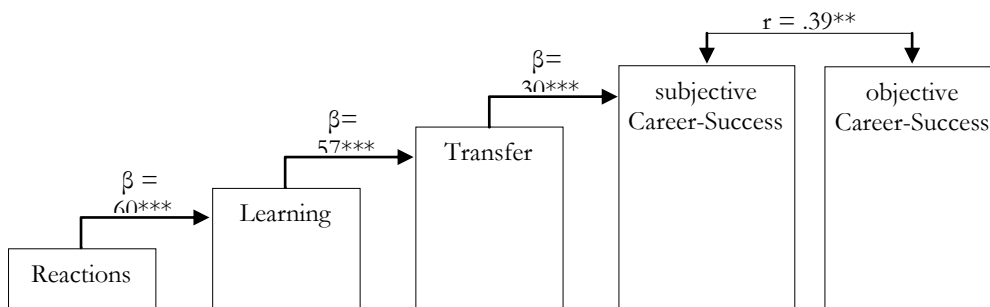
Notes:  $R^2 = .16$  for Step 1,  $\Delta R^2 = .16$  for Step 2; \*\*\* $p < .001$

The linear single regression with *transfer*-level as the independent and *subjective career-success* as the dependent variable shows that *transfer* with  $\beta = .30$  ( $p < .01$ ) is a predictor for *subjective career-success*, however, only 8.0 % of the variance is explained.

Due to the unclear research situation regarding the relationship between *subjective* and *objective career-success* (see Chapter 2.1.4), it is not possible to assign the factors to predictor and criterion variables. Although the correlation studies can be used to demonstrate a medium relationship between the two variables (see Table 1), no statement can be made about its modeling.

The results thus show that each previous level has an influence on the next level, albeit a weak one. Solely between the *transfer*-level and the *objective career-success* no connection can be proven, as previously mentioned, (see Figure 2).

Figure 2 Relationships between levels



Annotations: \*\* $p < .01$ , \*\*\* $p < .001$

## 5 Conclusion

The results show that the levels *reactions*, *learning*, *transfer* and *career-success* are interdependent. The first evaluation level *reactions* has an influence on the second level *learning*, which in turn is able to predict the *transfer*. Looking at the explained variance, the *reactions*-level and the *learning*-level prove to be good predictors for the following levels.

Concerning the relationship between the first two evaluation levels, the findings of this study are contrary to most studies (e.g. Alliger & Janak, 1989; Alliger et al., 1997; Arthur, Tubré, Paul & Edens, 2003; Colquitt et al., 2000; Gessler & Sebe-Opfermann, 2011; Ruona et al., 2002) as we identify a correlation (for similar results see Saks and Burke, 2012). This is remarkable as the understanding of the levels is similar to that of Alliger et al. (1997), albeit more comprehensive and precise, especially regarding the concept of *learning*. Our operationalization of the level *reactions* is thus suitable for predicting (non-)successful learning.

The connection between the *learning*-level and the *transfer*-level, on the other hand, is in line with the majority of the previously mentioned studies (e.g. Alliger & Janak, 1989; Alliger et al., 1997; Colquitt et al., 2000; Gessler & Sebe-Opfermann, 2011; Saks & Burke, 2012). Thus it can be stated that – across several studies – the *learning*-level seems to be suitable for predicting the *transfer*-level.

Furthermore, the *transfer*-level seems to determine the *subjective career-success*. However, as this only explains 8% of the variance it is only a minor influence and *subjective career-success* seems to mainly be determined by other factors. A comparison with previous research results concerning this relationship is difficult as, according to the authors' knowledge, this study is the first to understand and operationalize the fourth evaluation level as *career-success* (although Praslova, 2010, proposes – among other possible criteria, such as service to society – measuring the alumni's career-success instead of results as well, she did not empirically assess this suggestion). The results indicate that the (re)interpretation of Kirkpatrick's model is capable of generating insightful data. Especially as it enables an evaluation on all four levels – even if the organizational consequences of a continuing training program cannot be investigated for various reasons.

Despite this, the revised understanding of levels requires further investigations in order to allow conclusive, well-founded statements. It must be clarified whether *transfer* can indeed only predict such a small proportion of subjective career-success. Furthermore, it must be evaluated, whether the *transfer*-level is not a predictor of *objective career-success* even though *objective* and *subjective career-success* correlate with each other.

Although the distinction between *career-success* as an objective and a subjective one is common (Abele & Spurk, 2009; Abele et al., 2011; Barthauer et al., 2016; Dette et al., 2004; Judge & Kammeyer-Mueller, 2007; Spurk & Abele, 2014; Weber, 2013), it may not be suitable for the evaluation of further training measures. Moreover, we follow e.g. Huber (2011) and Webb, Schumacker and Tilford (2017), who demands studies with a longitudinal design or mixed methods approaches in order to gain more meaningful insights.

Essentially, one must still acknowledge a lack of a theoretical basis for the entire evaluation model. It must also be highlighted, that the conclusions of the various studies can only be compared to a limited extent, as each researcher used different definitions (and thus different operationalizations) for the evaluation levels.

It is therefore all the more remarkable that Kirkpatrick's evaluation model, which is impressive due to its simplicity and was not postulated to be a model in the scientific sense, but rather intended to provide guidelines for carrying out conclusive evaluations (Kirkpatrick 1970, 1979, 1996a; Kraiger 2003), is still valid in its basic assumptions – even 60 years after its first publication.

## References

- Abele, A. E., & Hagmaier, T. (2011). *Berufliche Laufbahnentwicklung von Akademikerinnen und Akademikern der Universität Erlangen-Nürnberg (BELA-E). Siebte Erhebung der prospektiven Längsschnittstudie. Fragebogen und Grundausswertung* (Projektbericht No. 7). Erlangen: Friedrich-Alexander-Universität Erlangen Nürnberg.
- Abele, A. E., Hagmaier, T., & Spurk, D. (2015). Does Career Success Make You Happy? *The Mediating Role of Multiple Subjective Success Evaluations. Journal of Happiness Studies*, 16(4), 1–19.
- Abele, A. E., & Spurk, D. (2009). How do objective and subjective career success interrelate over time? *Journal of Occupational and Organizational Psychology*, 82(4), 803–824.

- Abele, A. E., Spurk, D., & Volmer, J. (2011). The construct of career success: measurement issues and an empirical example. *Zeitschrift Für ArbeitsmarktForschung*, 43(3), 195–206.
- Alliger, G. M., & Janak, E. A. (1989). Kirkpatrick's levels of training criteria: thirty years later. *Personnel Psychology*, 42, 331–342.
- Alliger, G. M., Tannenbaum, S. I., Bennett, W. J., Traver, H., & Shotland, A. (1997). A meta-analysis of the relations among training criteria. *Personnel Psychology*, 50, 341–358.
- Arthur, M. B., Khapova, S. N., & Wilderom, C. P. M. (2005). Career success in a boundaryless career world. *Journal of Organizational Behavior*, 26(2), 177–202.
- Arthur, W. J., Tubré, T., Paul, D. S., & Edens, P. S. (2003). Teaching Effectiveness: The relationship between reaction and learning evaluation criteria. *Educational Psychology*, 23(3), 275–285.
- Arthur, W., Jr., Bennett, W., Jr., 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(2), 234–245.
- Aryee, S., Chay, Y. W., & Tan, H. H. (1994). An Examination of the Antecedents of Subjective Career Success Among a Managerial Sample in Singapore. *Human Relations*, 47(5), 487–509.
- Atria, M., Reimann, R., & Spiel, C. (2006). Qualitätssicherung durch Evaluation. Die Bedeutung von Zielexplikation und evaluativer Haltung. In C. Steinebach (Ed.), *Handbuch Psychologische Beratung*. Stuttgart: Klett-Cotta.
- Baldwin, T. T., Ford, J. K., & Blume, B. D. (2009). Transfer of Training 1988-2008: An Updated Review and Agenda for Future Research. In G. P. Hodgkinson & J. K. Ford (Eds.), *International Review of Industrial and Organizational Psychology* (pp. 41–70). Chichester, West Sussex, UK: John Wiley & Sons.
- Baron, R. M., & Kenny, D. A. (1986). The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations. *Journal of Personality and Social Psychology*, 51(6), 1173–1182.
- Barthauer, L., Estel, V., Dubbel, A., Kauffeld, S., & Spurk, D. (2016). Woran erkenne ich eine erfolgreiche Laufbahn? Ein qualitativer Ansatz zur Definition von Laufbahnerfolg bei Wissenschaftlern. *Beiträge Zur Hochschulforschung*, 38(1–2), 42–63.
- Bartz, D. E. (2019). Organization Members Directing their Career Development. *Journal of Education and Human Development*, 8(1), 43–47.
- Bates, R. (2004). A critical analysis of evaluation practice: the Kirkpatrick model and the principle of beneficence. *Evaluation and Program Planning*, 27(3), 341–347.
- Blankenship, R. L. (1973). Organizational Careers: An Interactionist Perspective. *The Sociological Quarterly*, 14(1), 88–98.
- Boudreau, J. W., Boswell, W. R., & Judge, T. A. (2001). Effects of Personality on Executive Career Success in the United States and Europe. *Journal of Vocational Behavior*, 58(1), 53–81
- Braun, E., Gusy, B., Leidner, B., & Hannover, B. (2008). Das Berliner Evaluationsinstrument für selbsteingeschätzte, studentische Kompetenzen (BEvaKomp). *Diagnostica*, 54(1), 30–42.
- Campion, M. A., & Campion, J. E. (1987). Evaluation of an interviewee skills training program in a natural field experiment. *Personnel Psychology*, 40, 675–691.
- 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.
- Day, E. A., Arthur, W., Jr., & Gettman, D. (2001). Knowledge structures and the acquisition of a complex skill. *Journal of Applied Psychology*, 86(5), 1022–1033.
- Detle, D. E., Abele, A. E., & Renner, O. (2004). Zur Definition und Messung von Berufserfolg: Theoretische Überlegungen und metaanalytische Befunde zum Zusammenhang von externen und internen Laufbahnerfolgsmäßen. *Zeitschrift für Personalpsychologie*, 3(4), 170–183.
- Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). Thousand Oaks, CA: SAGE Publications.
- Gegenfurtner, A. (2013). Dimensions of Motivation to Transfer: A Longitudinal Analysis of Their Influence on Retention, Transfer, and Attitude Change. *Vocations and Learning*, 6(2), 187–205
- Gessler, M. (2005). Gestaltungsorientierte Evaluation und der Return on Investment von Weiterbildungsprogrammen. *Bwp@ Berufs- Und Wirtschaftspädagogik - Online*, 9, 1–25.
- Gessler, M., & Sebe-Opfermann, A. (2011). Der Mythos “Wirkungskette” in der Weiterbildungsforschung - empirische Prüfung der Wirkannahmen im “Four Levels Evaluation Model” von Donald Kirkpatrick. *Zeitschrift Für Berufs- Und Wirtschaftspädagogik*, 107(2), 270–279.
- Gläßer, E., Gollwitzer, M., Kranz, D., Meiniger, C., Schlotz, W., Schnell, T., & Voß, A. (2002). TRIL - Trierer Inventar zur Lehrevaluation. Kurznachweis. Universität Trier.
- Göhlich, M., & Zirfas, J. (2007). Der pädagogische Grundbegriff des Lernens. *Odgojne Znanosti*, 9(2), 7–24.

- Greenhaus, J. H., Parasuraman, S., & Wormley, W. M. (1990). Effects of Race on Organizational Experiences, Job Performance Evaluations, and Career Outcomes. *The Academy of Management Journal*, 33(1), 64–86.
- Grohmann, A., & Kauffeld, S. (2013). Evaluating training programs: Development and correlates of the Questionnaire for Professional Training Evaluation. *International Journal of Training and Development*, 17(2), 135–155.
- Gunz, H. P., & Heslin, P. A. (2005). Reconceptualizing career success. *Journal of Organizational Behavior*, 26(2), 105–111.
- Heslin, P. A. (2005). Conceptualizing and evaluating career success. *Journal of Organizational Behavior*, 26(2), 113–136.
- Hogan, R. L., Cepela, D., & Fentress, B. (2014). Training Evaluation: Audience Resonse System as an Evaluation Tool. *Journal of Education and Human Development*, 3(2), 259–269.
- Holton III, E. F. (1996). The flawed four-level evaluation model. *Human Resource Development Quarterly*, 7(1), 5–21.
- Holton III, E. F. (2005). Holton's Evaluation Model: New Evidence and Construct Elaborations. *Advances in Developing Human Resources*, 7(1), 37–54.
- Huber, S. G. (2011). The impact of professional development: A theoretical model for empirical research, evaluation, planning and conducting training and development programmes. *Professional Development in Education*, 37(5), 837–853.
- Hughes, E. C. (1937). Institutional Office and the Person. *American Journal of Sociology*, 43(3), 404–413.
- Judge, T. A., Higgins, C. A., Thoresen, C. J., & Barrick, M. R. (1999). The big five personality traits, general mental ability, and career success across the life span. *Personnel Psychology*, 52(3), 621–652.
- Judge, T. A., & Kammeyer-Mueller, J. D. (2007). Personality and career success. In H. P. Gunz & M. A. Peiperl (Eds.), *Handbook of career studies* (pp. 59–78). Thousand Oaks, CA: Sage Publications.
- Kauffeld, S., Bates, R., Holton III, E. F., & Müller, A. C. (2008). Das deutsche Lerntransfer-System-Inventar (GLTSI): psychometrische Überprüfung der deutschsprachigen Version. *Zeitschrift für Personalpsychologie*, 7(2), 50–69.
- Kirkpatrick, D. L. (1959a). Techniques for evaluating training programs: Part 1 - reaction. *Journal of the American Society of Training Directors*, 13(11), 3–9.
- Kirkpatrick, D. L. (1959b). Techniques for evaluating training programs: Part 2 - learning. *Journal of the American Society of Training Directors*, 13(12), 21–26.
- Kirkpatrick, D. L. (1960a). Techniques for evaluating training programs: Part 3 - behavior. *Journal of the American Society of Training Directors*, 14(1), 13–18.
- Kirkpatrick, D. L. (1960b). Techniques for evaluating training programs: Part 4 - results. *Journal of the American Society of Training Directors*, 14(2), 28–32.
- Kirkpatrick, D. L. (1970). In *Evaluation of Training*. In *Evaluation of short-term training in rehabilitation*. Oregon studies in the rehabilitation of the retarded (pp. 35–61). Oregon: Department of Special Education, University of Oregon.
- Kirkpatrick, D. L. (1979). Techniques for Evaluating Training Programs. *Training and Development Journal*, 6, 78–92.
- Kirkpatrick, D. L. (1987). *More Evaluating Training Programs. A collection of articles from Training and Development Journal*. Alexandria: American Society for Training and Development.
- Kirkpatrick, D. L. (1996a). Evaluation. In R. L. Craig (Ed.), *The ASTD training and development handbook: a guide to human resource development* (4th ed., pp. 294–312). New York: McGraw-Hill.
- Kirkpatrick, D. L. (1996b). Great Ideas Revisited. Techniques for Evaluating Training Programs. Revisiting Kirkpatrick's Four-Level Model. *Training and Development*, 50(1), 54–59.
- Kirkpatrick, D. L. (1998). Great Ideas Revisited. In D. L. Kirkpatrick (Ed.), *Another look at evaluating training programs: fifty articles from Training & development and Technical training: magazines cover the essentials of evaluation and return-on-investment* (pp. 3–8). Alexandria, VA: American Society for Training & Development.
- Kirkpatrick, D. L., & Kirkpatrick, J. D. (2006). *Evaluating training programs: the four levels* (3rd ed.). San Francisco, CA: Berrett-Koehler.
- Kraiger, K. (2003). Perspectives on training and development. In W. C. Borman, D. R. Ilgen, & R. J. Klimoski (Eds.), *Handbook of psychology* (pp. 171–192). Hoboken: John Wiley & Sons.
- Kraiger, K., Ford, K. J., & 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.
- Machin, M. A., & Fogarty, G. J. (2004). Assessing the antecedents of transfer intentions in a training context. *International Journal of Training and Development*, 8(3), 222–236.
- Marsh, H. W. (1982). SEEQ: A Reliable, Valid, and Useful Instrument for Collecting Students' Evaluations of University Teaching. *British Journal of Educational Psychology*, 52(1), 77–95.

- Ng, T. W. H., Eby, L. T., Sorensen, K. L., & Feldman, D. C. (2005). Predictors of objective and subjective career success: a meta-analysis. *Personnel Psychology*, *58*, 367–408.
- Ng, T. W. H., & Feldman, D. C. (2014). Subjective career success: A meta-analytic review. *Journal of Vocational Behavior*, *85*(2), 169–179.
- Praslova, L. (2010). Adaptation of Kirkpatrick's four level model of training criteria to assessment of learning outcomes and program evaluation in Higher Education. *Educational Assessment, Evaluation and Accountability*, *22*(3), 215–225.
- Reio, T. G., Rocco, T. S., Smith, D. H., & Chang, E. (2017). A Critique of Kirkpatrick's Evaluation Model. *New Horizons in Adult Education & Human Resource Development*, *29*(2), 35–53.
- Rindermann, H. (2009). Qualitätsentwicklung in der Hochschullehre. *Beiträge zur Lehrerbildung*, *27*(1), 64–73.
- Ritzmann, S., Hagemann, V., & Kluge, A. (2014). The Training Evaluation Inventory (TEI) - Evaluation of Training Design and Measurement of Training Outcomes for Predicting Training Success. *Vocations and Learning*, *7*(1), 41–73.
- Rudolf, M., & Müller, J. (2012). *Multivariate Verfahren: eine praxisorientierte Einführung mit Anwendungsbeispielen in SPSS* (2nd ed.). Göttingen: Hogrefe.
- Ruona, W. E. A., Leimbach, M., Holton III, E. F., & Bates, R. (2002). The relationship between learner utility reactions and predicted learning transfer among trainees. *International Journal of Training and Development*, *6*(4), 218–228.
- Saks, A. M., & Burke, L. A. (2012). An investigation into the relationship between training evaluation and the transfer of training: Training evaluation. *International Journal of Training and Development*, *16*(2), 118–127.
- Salas, E., & Cannon-Bowers, J. A. (2001). The science of training: a decade of progress. *Annual Review of Psychology*, *52*, 471–499.
- Seibert, S. E., & Kraimer, M. L. (2001). The Five-Factor Model of Personality and Career Success. *Journal of Vocational Behavior*, *58*(1), 1–21.
- Spurk, D., & Abele, A. E. (2014). Synchronous and time-lagged effects between occupational self-efficacy and objective and subjective career success: Findings from a four-wave and 9-year longitudinal study. *Journal of Vocational Behavior*, *84*(2), 119–132.
- Spurk, D., Abele, A. E., & Volmer, J. (2011). The Career Satisfaction Scale: Longitudinal measurement invariance and latent growth analysis: Career Satisfaction Scale. *Journal of Occupational and Organizational Psychology*, *84*(2), 315–326.
- Super, D. E. (1980). A Life-Span, Life-Space Approach to Career Development. *Journal of Vocational Behavior*, *16*, 282–298.
- Tirre, S. (2012). *Wirksamkeit von Weiterbildung. Evaluation der Ausbildung zum Innovationspromotor*. Humboldt-Universität, Berlin.
- Warr, P., Allan, C., & Birdi, K. (1999). Predicting three levels of training outcome. *Journal of Occupational and Organizational Psychology*, *72*, 351–375.
- Warr, P., & Bunce, D. (1995). Trainee characteristics and the outcomes of open learning. *Personnel Psychology*, *48*, 347–375.
- Webb, A. L., Schumacker, R. E., & Tilford, A. (2017). Synthesis of Published Articles from Studies in Educational Evaluation, 2010-2015. *Journal of Education and Human Development*, *6*(2), 78–81.
- Weber, A. (2013). *Berufserfolg und Lebenszufriedenheit*. Universität Duisburg-Essen, Duisburg-Essen.
- Xiao, J. (1996). The Relationship Between Organizational Factors and the Transfer of Training in the Electronics Industry in Shenzhen, China. *Human Resource Development Quarterly*, *7*(1), 55–73.