Career Construction Across the Life Span: Career Choice and Career Development

Inaugural-Dissertation

zur Erlangung der Doktorwürde der

Fakultät für Humanwissenschaften

der

Julius-Maximilians-Universität Würzburg

vorgelegt von

Andrea Beinicke

aus Würzburg

2015

Erstgutachter: Professor Dr. Wolfgang Schneider

Zweitgutachter: PD Dr. Wolfgang Lenhard

Tag des Kolloquiums: 13.08.2015

List of abbreviations

ACT	American College Testing
AERA	American Educational Research Association
APA	American Psychological Association
ASTD	American Society for Training and Development
CFA	confirmatory factor analysis
CHC	Cattell-Horn-Carroll
DIF	differential item functioning
GIST-R	Revised General Interest Structure Test
HR	human resource
HRD	human resource development
IRT	item response theory
LTSI	Learning Transfer System Inventory
MASEM	meta-analytic structural equation modeling
MMB	MMB-Institute for Media and Competence Research
MMR	moderated multiple regression
NCME	National Council on Measurement in Education
NIE	National Institute of Education
Q4TE	Questionnaire for Professional Training Evaluation
RIASEC	Realistic, Investigative, Artistic, Social, Enterprising, Conventional
ROI	return on investment
RTOR	randomization tests of hypothesized order relations
SCCT	social cognitive career theory
SCT	social cognitive theory
SDS	Self-Directed Search
SII	Strong Interest Inventory
T&D	training and development
TIE	typical intellectual engagement
TWA	theory of work adjustment
UNESCO	United Nations Educational, Scientific and Cultural Organization
UNIACT-R	Revised Unisex Edition of the American College Testing

Summary

This dissertation contributes to deepen our understanding of constructs that play a key role in individuals' vocational career construction. In this regard, many previous studies have focused exclusively on a specific phase of an individual's career. Yet, modern societies require continuous investments in one's career to adapt to changing environments throughout the life span. Consequently, this dissertation takes a broad approach to capture a wide spectrum of career construction processes.

According to Super's (1990) developmental stage framework, individuals have to manage vocational developmental tasks corresponding to each of the developmental life stages in order to be career mature across the life span. As the two stages *exploration* and *maintenance* set the stage for individuals' future career pathways, they are especially important in individuals' vocational career construction. Therefore, both of them are addressed in this dissertation.

In the exploration stage, individuals have to make vocational choices in order to strive for career direction. Such vocational choices are guided by vocational interests in this early career stage, which provide the basis for individuals' career construction in the following career stages. This dissertation provides empirical evidence from three studies on vocational interests (Studies 1, 2, and 3). All of these studies aimed at inform the ongoing debate of (a) how vocational interests and specific cognitive abilities are related and (b) how interest inventories can suspend gender differences in instrument validity and prediction without losing structural validity.

The first study meta-analytically investigated the nature and magnitude of the relationship between vocational interests and both general as well as specific cognitive abilities. This comprehensive quantitative analysis demonstrated meaningful relationships between cognitive abilities and vocational interests. Meta-analytic coefficients' strength and direction were comparable for female and male samples. In contrast, age and birth cohort were detected as moderators of the relationship between vocational interests and cognitive abilities.

In response to NIE and APA Standards (NIE, 1975) that strongly emphasize the elimination of gender differences in instrument validity and prediction, the second study worked towards making vocational interest tests gender-fair. In this respect, gender-related

differential validity and differential prediction were assessed jointly with a differential item functioning (DIF) analysis for the first time together in the vocational interest domain as one major concern of test fairness. More specifically, in order to investigate potential prediction bias in vocational interest measures, gender-specific validity coefficients for the prediction of person–environment fit and satisfaction to test for differential validity were compared first. Second, gender differences in the slopes and intercepts of the regression model predicting person-environment fit were examined to test for differential prediction. Results showed evidence of differential validity and some indications of differential prediction in a standard Holland interest inventory. However, removing items showing large gender specific DIF slightly reduced prediction bias, thus, controlling for measurement bias.

Following up on these results, the third study provided insight into the question whether or not structural properties of a standard Holland interest inventory are affected when items showing gender-specific DIF are eliminated during the test construction process. Holland’s hexagonal model was tested for structural invariance using a confirmatory methodological approach (i.e., confirmatory factor analysis and randomization tests of hypothesized order relations). Results suggested that eliminating items showing gender-specific DIF had no considerable influence on the instrument’s psychometric structure. Overall, the results from studies on vocational interests emphasize the importance that (a) vocational interests should be considered in accordance with cognitive abilities and that (b) DIF is one possibility to significantly improve test fairness when developing interest inventories.

In the maintenance stage, on the contrary, individuals must strive for career adaptability in order to respond to rapidly changing labor needs of post-modern societies, once an occupation has been chosen, to maintain career success. Therefore, individuals focus on active career development. A crucial building block of career development is acquiring new knowledge, skills, or abilities through the successful completion of trainings. In this regard, guaranteeing training success is important to ensure that individuals make satisfactory progress in their career development. In this regard, this dissertation provides empirical evidence from three studies on training evaluation (Studies 4, 5, and 6). All of these studies aimed at investigating the transfer problem in the workplace, especially in corporate e-learning settings.

The fourth study explored the prevalence of transfer support actions in companies which are diverse in branches of industry and among the best employers in Germany to investigate the magnitude of the gap between theory and practice (theory-practice-transfer). Results showed that evidence-based recommendations and best practices for maximizing training effectiveness were often implemented in both corporate e-learning and classroom training settings before, during, and after training. However, transfer support actions were least implemented after training compared to before and during training. Comparing both settings, such actions were more prevalent in classroom trainings than in e-learning settings across all of the time periods.

The fifth study systematically compared objective and subjective training success across corporate e-learning and classroom training. Therefore, the study contributed to experimental training effectiveness studies with a specific focus on e-learning settings. Results revealed that the scores in the overall performance test did not vary across training settings per se, but did so when type of knowledge and timing of the assessment were taken into account. More specifically, factual knowledge was more effectively learned in e-learning settings, whereas applied knowledge was more effectively trained in classroom training settings when assessing knowledge immediately after training. However, no difference occurred across training settings six to eight weeks after training, making the initial differences rather volatile. Subjective success of learning and success of transfer did not vary across training settings either. Moreover, high correlations between objective and subjective training success measures were found.

Finally, the sixth study explored which transfer factors are substantial for training success in corporate e-learning settings. Furthermore, the mediating role of success of learning in the relationship between transfer factors and success of transfer was explored in separate path models for corporate e-learning and classroom settings using a meta-analytic structural equation modeling technique. Results identified transfer motivation as the most crucial variable to ensure success of learning and success of transfer in corporate e-learning environments. Overall, the results from all of the three studies on training evaluation emphasize the importance of promoting variable, training specific transfer factors in corporate training and transfer environments so that time, money, and resources of an organization are not wasted.

By answering open research questions relevant to career choice in early career stages and to career development in later career stages, this dissertation contributes to the overarching goal of shedding more light on constructs relevant to individuals' vocational career construction processes across the life span. Beyond the results presented within each study's horizon, this dissertation aimed at offering practical guidance to career counselors, trainees, and training and development (T&D) professionals. Career counselors and T&D professionals are involved in guiding vocational career construction processes of individuals across the life span. Thus, on the one hand, this dissertation supports career counselors' work so that they can help deliberating individuals make optimal and effective career choices. On the other hand, this dissertation facilitates T&D professionals' work so that they can effectively design and evaluate e-learning and classroom trainings in corporate educational settings. Identifying individuals' vocational interests combined with cognitive abilities through adequate test measures and maximizing success of learning and success of transfer through fostering evidence-based transfer support actions will help individuals adapt quickly to the changing nature of work environments in the 21st century and to continue to successfully construct careers across the life span.

Zusammenfassung

Diese Promotion trägt dazu bei, unser Verständnis jener Konstrukte zu vertiefen, die eine entscheidende Rolle bei der beruflichen Karriereplanung von Individuen spielen. Viele bisherige Studien haben sich in dieser Hinsicht ausschließlich auf eine spezielle Karrierephase konzentriert. Jedoch verlangt die moderne Gesellschaft, dass Individuen kontinuierlich in ihre Karriere investieren, um sich über die Lebensspanne hinweg an die ständig verändernden Umweltbedingungen anpassen zu können. Demzufolge wählt diese Promotion einen umfassenden Ansatz mit dem Ziel, ein breiteres Spektrum an Prozessen der Karriereplanung zu erfassen.

Laut des Entwicklungsstufenkonzepts über die Lebensspanne von Super (1990) müssen Individuen in jeder der Entwicklungsstufen berufliche Entwicklungsaufgaben meistern, um der Karriere - über die Lebensspanne - gewachsen zu sein. Da die beiden Stufen *Exploration* und *Erhaltung* zukünftige Karrierewege bahnen, sind diese bei der beruflichen Karriereplanung von Individuen besonders wichtig und stehen deshalb im Fokus dieser Promotion.

Auf der Explorationsstufe müssen Individuen berufliche Entscheidungen treffen, um eine Karriererichtung einschlagen zu können. Solche beruflichen Entscheidungen in dieser frühen Karrierestufe werden von beruflichen Interessen geleitet, welche die Grundlage für die weitere individuelle Karriereplanung legen. Diese Promotion liefert empirische Befunde aus drei Studien zum Thema berufliche Interessen (Studien 1, 2 und 3). Alle diese Studien intendieren, die andauernde Diskussion bezüglich der Fragen anzuregen: Wie hängen berufliche Interessen und spezifische kognitive Fähigkeiten miteinander zusammen? Auf welche Weise können Geschlechterunterschiede bezüglich instrumenteller Validität und Vorhersage in Interesseninventaren eliminiert werden, ohne strukturelle Validität zu verlieren?

Die erste Studie untersuchte metaanalytisch die Art und die Höhe des Zusammenhangs zwischen beruflichen Interessen und sowohl allgemeinen als auch spezifischen kognitiven Fähigkeiten. Die umfassende quantitative Analyse zeigte bedeutende Zusammenhänge zwischen kognitiven Fähigkeiten und beruflichen Interessen. Die metaanalytisch berechneten Koeffizienten waren in Stärke und Richtung für Männer und

Frauen vergleichbar. Alter und Geburtsjahrgang wurden als Moderatoren des Zusammenhangs zwischen beruflichen Interessen und kognitiven Fähigkeiten identifiziert.

Gemäß der Standards von NIE und APA (NIE, 1975), welche die Eliminierung von Geschlechterunterschieden in instrumenteller Validität und Vorhersage stark betonen, beschäftigte sich die zweite Studie damit, berufliche Interessentests den Geschlechtern entsprechend gerecht zu konstruieren. Diesbezüglich wurden die geschlechtsabhängige differentielle Validität und die differentielle Vorhersage erstmalig im Zusammenhang mit differential item functioning (DIF) Analysen im Bereich beruflicher Interessen als einen Ansatz für Testfairness untersucht. Zur Ermittlung möglicher Verzerrungen in den Vorhersagen aus beruflichen Interessentests wurden erstens geschlechtsspezifische Validitätskoeffizienten für die Vorhersage der Person-Umwelt-Passung und der Zufriedenheit miteinander verglichen, um die differentielle Validität zu prüfen, und zweitens Geschlechterunterschiede in den Steigungen und Achsenabschnitten des Regressionsmodells zur Vorhersage der Person-Umwelt-Passung untersucht, um die differentielle Vorhersage zu prüfen. Die Ergebnisse lieferten Belege für die differentielle Validität und Hinweise auf differentielle Vorhersage in einem üblichen Holland Interesseninventar. Nach der Entfernung von Items, die ein großes geschlechtsspezifisches DIF zeigten, reduzierten sich jedoch die Verzerrungen in der Vorhersage etwas. Die Anwendung von DIF erlaubt die Kontrolle dieser Verzerrungen.

Die dritte Studie beleuchtete die Frage, inwieweit sich das Eliminieren von Items mit geschlechtsspezifischen DIF während der Testkonstruktion auf die psychometrische Struktur eines üblichen Holland Interesseninventars auswirkt. Holland's Hexagonmodell wurde auf strukturelle Invarianz getestet. Hierzu wurde ein konfirmatorischer Ansatz gewählt (konfirmatorische Faktorenanalyse und randomization tests of hypthesized order relations). Die Ergebnisse deuten darauf hin, dass das Eliminieren von Items mit geschlechtsspezifischen DIF keinen nennenswerten Einfluss auf die psychometrische Struktur des Instruments hat. Es wird diskutiert, inwiefern die Verwendung von DIF eine Möglichkeit zur Verbesserung der Testfairness bei der Entwicklung von Interesseninventaren darstellt. Insgesamt unterstreichen die Ergebnisse dieser drei Studien zum Thema beruflicher Interessen, wie wichtig es ist, (a) berufliche Interessen zusammen mit kognitiven Fähigkeiten zu berücksichtigen und (b), dass DIF eine Möglichkeit darstellt, um die Testfairness bei der Konstruktion von Interesseninventaren signifikant zu erhöhen.

Nach ihrer Berufswahl, auf der Stufe der Erhaltung, müssen Individuen nach karrierebezogener Anpassungsfähigkeit streben, um auf die sich schnell verändernden Anforderungen der postmodernen Arbeitswelt reagieren zu können. Zu diesem Zweck konzentrieren sie sich aktiv auf ihre Karriereentwicklung. Einen wichtigen Baustein der Karriereentwicklung bildet die Aneignung von neuem Wissen, Kompetenzen und Fertigkeiten im Rahmen von erfolgreichen Trainings. In dieser Hinsicht ist es wichtig, den Trainingserfolg sicherzustellen und damit den Individuen zufriedenstellende Fortschritte in ihrer Karriereentwicklung zu ermöglichen. Diese Promotion liefert empirische Befunde aus drei Studien zum Thema Trainingsevaluation (Studien 4, 5 und 6). Alle diese Studien hatten zum Ziel, das Transferproblem am Arbeitsplatz, besonders im Rahmen von E-Learning in Unternehmen, zu untersuchen.

Die vierte Studie untersuchte die Verbreitung von transferförderlichen Maßnahmen in Unternehmen aus verschiedenen Industriebranchen, die zu den besten Arbeitgebern in Deutschland zählen, mit dem Ziel, die Diskrepanz zwischen Theorie und Praxis (Theorie-Praxis-Transfer) genauer zu beschreiben. Die Ergebnisse veranschaulichen, dass evidenzbasierte Empfehlungen und Techniken zur Optimierung der Trainingseffektivität bei E-Learning und Präsenztrainings sowohl vor, während, als auch nach dem Training sehr häufig angewendet wurden. Jedoch zeigte sich, dass transferförderliche Maßnahmen nach dem Training, verglichen mit während und vor dem Training, am wenigsten eingesetzt wurden. Beim Vergleich beider Trainingsformen wurde deutlich, dass diese Maßnahmen bei Präsenztrainings im Vergleich zum E-Learning über alle gemessenen Zeitpunkte häufiger verbreitet waren.

Die fünfte Studie stellte objektive und subjektive Maße von Trainingserfolg für E-Learning und Präsenztrainings in Unternehmen systematisch gegenüber. Damit leistete die Studie einen Beitrag zur experimentellen Erforschung der Trainingseffektivität, mit speziellem Fokus auf E-Learning. Die Ergebnisse zeigten, dass sich die Werte des Leistungstests insgesamt nicht zwischen den Trainingsformen per sé, aber zwischen den Wissensformen und Zeitpunkten der Erhebung unterschieden. Unmittelbar nach dem Training wurde Faktenwissen effektiver durch E-Learning vermittelt, während Anwendungswissen effektiver durch Präsenztraining vermittelt wurde. Allerding wurde kein Unterschied beim Anwendungswissen zwischen den Trainingsformen sechs bis acht Wochen später gefunden, was darauf hindeutet, dass sich die anfänglichen Unterschiede aufheben. Ferner zeigte sich kein Unterschied der Trainingsformen im subjektiven Lern- und Transfererfolg. Zusätzlich wurden hohe Korrelationen zwischen objektiven und subjektiven Maßen des Trainingserfolgs gefunden.

Die sechste Studie untersuchte explorativ die Bedeutung verschiedener Transferfaktoren für den Trainingserfolg in E-Learning-Umgebungen von Unternehmen. Außerdem wurde die Mediatorfunktion von Lernerfolg für den Zusammenhang zwischen Transferfaktoren und Transfererfolg in separaten Pfadmodellen jeweils für E-Learning und Präsenztrainings mittels meta-analytischen Strukturgleichungsmodellen untersucht. Die Ergebnisse ermittelten Transfermotivation als den entscheidenden Schlüssel zur Sicherung des Lern- und Transfererfolgs in unternehmensbezogenen E-Learning-Umgebungen. Insgesamt unterstreichen die Ergebnisse dieser drei Studien zum Thema Trainingsevaluation die Bedeutung der Förderung von variablen trainingsspezifischen Transferfaktoren für das Training in Unternehmen und die Transferumgebung, um Zeit, Geld und Ressourcen des Unternehmens sinnvoll einzusetzen.

Durch die Beantwortung offener Forschungsfragen, die sowohl für die Berufswahl in frühen Karrierestufen als auch für die Karriereentwicklung in späteren Karrierestufen relevant sind, trägt diese Promotion zu dem übergeordneten Ziel bei, Konstrukte zu beleuchten, die für den Prozess der individuellen Karriereplanung über die Lebensspanne bedeutsam sind. Neben den Ergebnissen, die im Rahmen der Studien präsentiert wurden, beabsichtigt diese Promotion Karriereberatern, Lernenden und Personalentwicklern praktische Hilfestellungen zu geben. Karriereberater und Personalentwickler sind an der Lenkung von Prozessen der beruflichen Karriereplanung über die Lebensspanne beteiligt. Deshalb versucht diese Promotion einerseits die Arbeit von Karriereberatern zu unterstützen, damit sie unentschlossenen Individuen dabei helfen können, optimale und effektive Karriereentscheidungen zu treffen. Andererseits versucht diese Promotion Personalentwicklern Unterstützung zu bieten, damit sie Umgebungen für E-Learning und Präsenztrainings im Unternehmen effektiv gestalten und bewerten können. Die Identifikation individueller beruflicher Interessen zusammen mit kognitiven Fähigkeiten mittels adäquater Testverfahren und die Maximierung des Lern- und Transfererfolgs mittels Implementierung evidenzbasierter transferförderlicher Maßnahmen sollen dabei helfen, sich schnell an die Veränderungen der Arbeitsumgebungen des 21. Jahrhunderts anzupassen, und eine erfolgreiche Karriere über die Lebensspanne zu durchlaufen.

CHAPTER 1 INTRODUCTION TO CAREER CONSTRUCTION ACROSS THE LIFE SPAN

Individuals seek for giving meanings und purposes to their journeys of life. Individuals write their individual career stories by attaching value to their previous life experiences. Especially in the post-modern world, individuals endeavor to exploit their full potential by investing in a successful career. Due to the fact that individuals increasingly focus on their career development across the life span, career-enhancing issues are of great interest for everybody in the work environment.

Career development and counseling have developed a strong theoretical and empirical base in the last few decades (Leung, 2008). Drawn from the research literature in the U.S., theories and frameworks of career development have been internationally recognized, empirically proven, and guided career counseling research and practice (Brown & Lent, 2004; J. L. Swanson & Gore, 2000).

Traditional theories of career choice and career development

Derived from the historical process in time, traditional theories of career choice and career development are subdivided into three categories: (a) *person-environment fit*, (b) *developmental*, and (c) *social cognitive* (Savickas, 2013). For an overview, these concepts are briefly summarized in the following.

First, in the mid and late 20th century, theoretical models focused on *person-environment fit*, that is, how to best match an employee to the work environment. Especially, the Minnesota theory of work adjustment (TWA; Dawis, 2002, 2005; Dawis & Lofquist, 1984) as well as Holland’s theory of vocational personalities in work environments (J. L. Holland, 1985, 1997) are well-investigated theories that similarly emphasize the person-environment correspondence (Dawis, 1992). In both theories, high levels of congruence between the individual’s characteristics (Holland: personality-interest type) and the work environment most likely lead to vocational satisfaction and stability and vice versa. Following Holland’s theory, a person’s vocational behavior is determined by an interaction between his or her personality and the characteristics of his or her environment such as workplace or studies. Thus, the individual’s interest profile that corresponds to the interest profile of the environment leads to Holland's hexagonal RIASEC model. The concept of Holland's six typologies (*R* Realistic, *I* Investigative, *A* Artistic, *S* Social, *E* Enterprising, and *C* Conventional) results in an individual three-letter code that directs career choice and satisfaction.

Second, in contrast to person-environment fit theories, *developmental* theories consider the process of choosing a career across the life span. Thus, contextual factors, such as the interference of work and other life roles, receive more attention. Following the developmental stage framework across the life span proposed by Super (1990), individuals have to manage vocational developmental tasks corresponding to each of the developmental life stages (growth, exploration, establishment, maintenance (or management), and disengagement) to be career mature. In the course of developmental theories, L. S. Gottfredson's theory of circumscription and compromise (1981, 2004) also incorporates the availability of career choices in individuals' career development depending on gender roles and social status.

Third, from a *social cognitive* approach, social learning theory (Mitchell & Krumboltz, 1996) assumes that self-constructed beliefs about oneself derive from the valence of received feedback. Positive feedback towards a certain behavior enhances the possibility to maintain interest in that specific activity. The social cognitive approach subsumes the dynamic interaction between individuals (e.g., gender), environment (e.g., support), and behavior. Following this perspective, a person is seen as an agent who actively influences his or her own environment by showing a certain behavior. On the basis of ensuing feedback, the individual forms specific cognitions about himself or herself and the surrounding environment. While Bandura's social cognitive theory (SCT; Bandura, 1986) focuses on the individual's interpretation of a specific event, the social cognitive career theory (SCCT; Lent, Brown, & Hackett, 1994) sets its focus on the individual's cognitive process and how this affects career-related beliefs and behavior. In 1994, the SCCT was first stated by Lent et al. in their article "Toward a unifying social cognitive theory of career and academic interest, choice and performance". Here, the SCCT serves as a framework for describing aspects (e.g., self-efficacy, outcome expectations, or personal goals) and their interplays that affect individuals' career development. The SCCT sequentially describes three major stages: (a) how individuals form interests that are relevant to their careers (that process is namely influenced by self-efficacy and outcome expectations), (b) how career choice options are selected and developed according to interests, and (c) how these choices are put into action and performed in vocational tasks (whether they fail or succeed). The SCCT is noted as an emerging theory of career choice and development and thus bridges traditional and new approaches.

Critical review of traditional theories leads to a new approach

To explain the variety of differences in individuals' experiences in today's work environment, traditional vocational theories are essential and helpful. That is, person-environment fit theories emphasizing individual differences (Dawis, 1992), and developmental theories with the focus on a free development of the individual resulting in diverse interactions between the individual's self and vocational choices. SCCT connects traditional theories and new approaches because individual variables (e.g., self-efficacy and outcome expectations) and environmental variables (e.g., background contextual affordances) are interrelated in one model (J. L. Swanson, 2013). From a counseling point of view, these traditional theories remain essential since they explain individuals' career choices in early career stages. Thus, dealing with points in time of choice (e.g., career counseling for choosing a specific academic major in seeking higher education at university), these traditional theories are well applicable. Contributions of traditional theories serve as a solid foundation to more recent approaches (Blustein, 2006; Savickas, 2011).

However, it is commonly criticized that traditional theories in vocational psychology do not consider changes of individuals' work and life roles or changes of economic and labor market factors (J. L. Swanson, 2013). Undoubtedly, these structural changes tremendously affect individuals' career experiences across time.

The changing nature of work in the 21st century: flexibility instead of stability

In the 20th century, individuals' career pathways were predetermined by stability of work environments; that is, employees aimed at lifelong job positions and in return employers kept their employees close and guaranteed job security. Consequently, individuals only had to choose their occupation once in a life-time to aim at career success. Explaining career development using only person-environment models and vocational development models emphasizing commitment and stability seemed insufficient (Leung, 2008). However, the assumption of life-time commitment to one employer does not hold any longer.

At the beginning of the 21st century, we face a new social arrangement of work: Driven by the globalization and by information technologies that rapidly improve, career choices and possibilities seem far less definable and predictable (Savickas et al., 2009). Instability in businesses arises from mergers and acquisitions, and this leads to an increasing need for sufficient flexibility in the unpredictable future.

In addition, the change of work environment is characterized by a shift in the gender ratio of employees. In the mid-20th century, the work environment was primarily dominated by men, whereas family and home belonged to women. Across the course of time, traditional gender stereotyping has become blurred. As more and more women seek to fulfill themselves not only in the family domain but also in terms of vocational career, women take on dual life roles. The shift is supported additionally by men's changing life roles as displayed in the following example. After childbirth more and more women edge earlier into the labor market than in the past, and men increasingly fill in family roles by staying at home and taking care of children, too. As a natural consequence, altogether more women are employed partially and on a full-time basis compared to previous decades. This issue is discussed in a new approach of career theories called ecological model of women's career development that accounts for women's multiple life roles and responsibilities (Cook, Heppner, & O'Brien, 2002). We know from various research studies that gender differences are scientifically proven, for example, in leadership styles (Collins & Singh, 2006; Paustian-Underdahl, 2012) or teamwork processes (Ivanova-Stenzel & Kübler, 2011). Gender diversity in work teams has revealed an increasing body of research literature (Biernat & Sesko, 2013; Milliken & Martins, 1996; Van Knippenberg & Schippers, 2007). Thus, these discrepancies in employed women and men–resulting from a different understanding of life roles and different life experiences in women compared to men–must be appropriately taken into consideration in the 21st century labor market.

Looking at statistics demonstrating change over time, job transitions have become more and more frequent. Between the ages of 18 to 36 we change the job 9.6 times and take new jobs with new opportunities (U.S. Department of Labor, Bureau of Labor Statistics, 2002). In Germany, a current study by the Forsa Institute on behalf of XING–the largest professional network in the German-speaking area–revealed that every third (35%) German employee is willing to change the job; including 7% that already planned their job transition very concretely (XING, 2013, para. 1). Instead of finding stability on the labor market (congruence between education and occupation), workers are required to react flexible to maintain employability. But how should individuals respond to the unstable and changing environment theoretically? To answer this question, a new approach to career theories emerged in the debate about modern and post-modern career theory, research, and practice: career construction theory (Savickas, 2005).

Career construction theory approach to career construction processes

Career construction theory aims at describing vocational behavior across the life-cycle (Savickas, 2005). Derived from Super's life-span, life-space theory (1990), career construction theory takes up the developmental aspect. Thus, it explains the processes through which "individuals construct themselves, impose direction on their vocational behavior, and make meaning of their careers" (Savickas, 2013, p. 1). The rationale of career construction theory illustrates *how* individuals construct their careers by using their vocational personality to adapt to the changing nature of work environment in the 21st century.

Two aspects are of great interest: First, *how* individuals make vocational choices and second, *how* individuals maintain successful and satisfying work lives (Savickas, 2006). Derived from both modern and post-modern career theories, career construction theory answers these questions by combining three different perspectives on vocational behavior: (a) differential, (b) developmental, and (c) dynamic.

(a) Individual-differences psychology: Theories of vocational personality types (J. L. Holland, 1997) focus on the question of *what* different individuals prefer to do. This aspect is described as differential.
(b) Developmental psychology: The process of psycho-social adaption explains *how* individuals cope with developmental tasks and changes in their vocational environment (Super, 1969, 1980, 1990).
(c) Narrative psychology: Career construction theory emphasizes the dynamic process of *why* individuals build so-called life themes. The theory examines how individuals impose personal meaning and direction on vocational behavior as the final goal. To achieve this goal, these individual patterns of experiences and ambitions for the vocational future form such life themes.

As Super's (1990) developmental stage framework across the life span plays a pioneering role here, career construction theory directly refers to two of Super's developmental life stages. As the two stages exploration and maintenance pioneer individuals' future career pathways, they are especially important in individuals' vocational career construction. Therefore, both of them are addressed in this dissertation.

Career construction in early career stages. In early career stages as in the exploration stage, individuals have to make vocational choices in order to strive for career direction. Thus, as the developmental task is characterized by studying, training or job seeking, individuals focus on their career choices. Such career choices are guided by vocational interests in this early career stage, which provide the basis for individuals' career construction in the following career stages.

Career construction in later career stages. In later career stages as in the maintenance stage, on the contrary, individuals must strive for career adaptability to respond to rapidly changing labor needs of post-modern societies once an occupation is chosen to maintain career success. Therefore, individuals focus on active career development. A crucial building block of career development is acquiring new knowledge, skills, or abilities through the successful completion of trainings. In this regard, guaranteeing training success is important to ensure that individuals make satisfactory progress towards their career development.

The doctoral thesis addresses those two developmental career stages (exploration and maintenance) which are especially important in individuals' vocational career construction processes. This dissertation contributes to the overarching goal of shedding more light on constructs relevant to individuals' vocational career construction processes across the life span. The chosen sequence of the studies is intended to follow the chronology in individuals' career construction processes.

1.1 Early career stages: from vocational interests to career choice

The vocational career developmental process starts with selecting an occupation. The varieties of vocational possibilities leave freedom to choose a career, especially when apprenticeship or study programs are financially supported by the government as it is the case, for example, in Germany. Instantly, the question arises: What are possible markers adolescents can rely on in the phase of exploring and choosing career-enhancing steps that lead to a successful career? Career development is primarily affected by intrapersonal (e.g., interests), social (e.g., family and school) and environmental-societal (e.g., community) influences (Patton & McMahon, 2006). Intrapersonal influences namely vocational interests as one key affecting career-related choice behavior in an early career stage are particularly further focused in the following.

Over the last 30 years, a wide variety of research has been conducted in the field of vocational interests due to its relevance in career assessment and vocational guidance. Research revealed a vast amount of insights, and recent meta-analyses provide further evidence for the relevance of vocational interests for job satisfaction (Spokane, Meir, & Catalano, 2000), employee performance (Nye, Su, Rounds, & Drasgow, 2012; Van Iddekinge, Roth, Putka, & Lanivich, 2011), and turnover (Van Iddekinge et al., 2011).

Vocational interests predict work-related behaviors that we favor; consequently, we tend to choose occupations that are in congruence with our interests. Person-environment congruence has been found to be a relevant condition of job satisfaction (Spokane et al., 2000). Meta-analytic results based on 60 studies and 568 correlations showed that interests are moderately correlated with performance and persistence in work and academic contexts (Nye et al., 2012). Moreover, congruence indices–quantifying the person-environment fit between individuals and their occupation–were found to stronger predict performance than interest scores alone. Another meta-analysis by Van Iddekinge et al. (2011) revealed that the criterion-related validity of vocational interest measures was strong and increased when interests mattered in the job performed.

1.2 Later career stages: from career development to training evaluation

As the story of vocational life continues, after individuals chose their profession (e.g., becoming a dentist), they follow their goals by investing in all kinds of educational training (e.g., studying dentistry at university). When individuals managed to complete job-specific educational training and are successfully placed within their career-specific field of interest (e.g., oral surgeon at a dental clinic), they have to cope with questions such as: How am I and will I be successful in the job, not only right now but also in the long term in the future?

To be successful in the individual's career construction process over a long-time period, it is insufficient to rely on formerly acquired skills only. As knowledge societies emerged, the role of knowledge development and lifelong learning continues to grow (United Nations Educational, Scientific and Cultural Organization [UNESCO], 2005). Investing in educational and further training confers competitive advantages. Meeting the needs of continuous innovation, information and communication technology provide tools to develop knowledge regardless of time and place. Lifelong learning raises the probability to be successful in the job and on the job market. Employees have to meet career-specific challenges requiring the development of (new) skills and competencies; thus, job-specific training is essential. As we constantly must adapt to changing work environments, the whole process of training continuously repeats itself.

According to statistics, it seems that companies respond to the changing nature of the work environment. Increasing evidence shows that more and more companies are investing in extensive human resource development (HRD) activities to survive the "war for talent". Looking at training from a financial perspective, the State of the Industry Report which is annually published by the Association for Talent Development, formerly American Society for Training and Development (ASTD), the world's largest association dedicated to the T&D profession, found that U.S. organizations spent $156.2 billion on employee learning and development in 2011 (ASTD, 2012, para. 2), up from $125.8 billion in 2009. Due to technical and economic advantages of web-based training compared to traditional face-to-face training, an increasing number of companies has shifted from traditional classroom training to e-learning in recent decades. E-learning is defined as "the use of electronic technologies to deliver information and facilitate the development of skills and knowledge" (ASTD, 2012, "Content distribution," para. 6). Technology-based training settings account for 37.3% of hours spent on formal training (ASTD, 2012, “Content distribution,” para. 6). In Germany, €28.6 billion was spent on training of employees in 2010 (Seyda & Werner, 2012) which describes a nominal growth of 6.4% compared to 2007 (Lenske & Werner, 2009) and 7.7% compared to 2010 (Kauffeld, 2010). The annual investment amounted to €1,035 per employee with declining direct costs such as participation fees. Although indirect costs (e.g., facilities and administrative costs) remained stable, the duration spent on further training has increased significantly. In addition, about every fourth company implemented computer-based learning referring to computer-based training, web-based training, and e-learning (Seyda & Werner, 2012).

Since lifelong learning is a key-motor for individuals’ career development, it seems relevant to investigate samples of the work force in the course of T&D. In sum, this soaring commitment to further training of employees highlights the importance of current and future training. Today, companies share the same opinion that competencies of employees are one of the key elements to gain a competitive advantage (Salas, Tannenbaum, Kraiger, & Smith-Jentsch, 2012). Thus, future success of companies is directly related to how quickly employees learn and subsequently apply to practice what they have learned (Kauffeld, 2010). Investments in training programs are only of economic advantage if the content learned is directly transferred to the workplace and thus contributes to job performance improvement. The concept of training success with regard to practice and empirical determinants is considered in chapter 4 and chapter 5 comprising later career stages: from career development to training evaluation. Thus, this dissertation pursues a scientific research interest to make training success measurable and to identify predictors as well as interacting variables.

EARLY CAREER STAGES: FROM VOCATIONAL INTERESTS TO CAREER CHOICE

According to the rationale of career construction theory, it is of great interest *how* individuals construct their careers by using their vocational personality (Savickas, 2006). In early career stages as in the exploration stage (Super, 1990), individuals must strive for career direction. Thus, as the developmental task is characterized by studying, training or job seeking, individuals focus on their career choices. Vocational personality in terms of making career choices is driven by individual interests. But how are vocational interests structured theoretically? How are vocational interests measured? Do we find individual differences, especially according to gender? These questions lead directly to the issue of vocational interests which is the focus of interest in the following.

Chapter 2 starts with an integrative literature review of vocational interests. First, well-investigated theoretical structural models of vocational interests are briefly described (section 2.1). Second, a review in light of the relationship between vocational interests and general as well as specific cognitive abilities is presented (section 2.2). Third, vocational interests and gender differences are displayed within the current research literature focusing on (a) the relationship between interests and cognitive abilities, and (b) gender differences in the measurement of vocational interest inventories which leads to the crucial aspect of construct validity (section 2.3).

Chapter 3 provides empirical evidence of three studies on vocational interests. The first study meta-analytically investigates the nature and magnitude of the relationship between vocational interests and both general as well as specific cognitive abilities. Furthermore, it is explored whether this relationship varies by gender, age, and birth cohort as possible moderators (section 3.1). The second study addresses the issue of gender-related differential validity and differential prediction in vocational interest inventories as one major concern of test fairness (section 3.2). The third study examines how gender-specific DIF affects the psychometric structure of a standard RIASEC interest inventory (section 3.3).

CHAPTER 2 LITERATURE REVIEW OF VOCATIONAL INTERESTS

2.1 Vocational interest theories

Definitions of vocational interests

According to Lowman (2010), interests can be defined as “relatively stable psychological characteristics of people [that] identify the personal evaluation … attached to particular groups of occupational or leisure activity clusters” (p. 477).

The purpose of this section 2.1 is to focus on the most cited and well-known theoretical structural models of vocational interests: (a) J. L. Holland’s (1985, 1997) hexagonal model (section 2.1.1) and (b) Gati’s (1979, 1991) hierarchical model including its alternative hierarchical model by Rounds and Tracey (1996) [section 2.1.2]. Both hexagonal as well as hierarchical models refer to Holland’s six personality types. As a conclusion (section 2.1.3), both types of models are compared with regard to their validity across culture and gender.

Chapter 2 is based on the following publications:

Beinicke, A., Pässler, K., & Hell, B. (2014). Does gender-specific differential item functioning affect the structure in vocational interest inventories? *International Journal for Educational and Vocational Guidance*, *14*(2), 181-198. doi:10.1007/s10775-013-9254-y
The final publication is available at link.springer.com.

Pässler, K., Beinicke, A., & Hell, B. (2014). Gender-related differential validity and differential prediction in interest inventories. *Journal of Career Assessment*, *22*(1), 138-152. doi:10.1177/1069072713492934

Pässler, K., Beinicke, A., & Hell, B. (2015). Interests and intelligence: A meta-analysis. *Intelligence*, *50*, 30-51. doi:10.1016/j.intell.2015.02.001

2.1.1 Hexagonal model by Holland

J. L. Holland's (1959, 1997) theory of vocational interests and career choices is the most prevalent taxonomy of vocational interests and has received robust empirical support. Holland supposed that individuals seek and enter environments that allow them to express their interests and values and exercise their abilities and skills. Satisfaction with educational and occupational choices as well as performance and persistence is determined by the degree of fit between the individual's interest type and environmental requirements. Holland's theory assumes that most individuals and environments can be categorized into one of six types: Realistic (R), Investigative (I), Artistic (A), Social (S), Enterprising (E), and Conventional (C), collectively referred to as RIASEC. Each type can be distinguished by specific interests, abilities, competencies, values, and life goals. According to Holland, the Realistic type prefers activities that involve working with tools, machines, or outdoor; the Investigative type shows a preference for sciences; the Artistic type is interested in the creative expression of ideas through writing or visual and performing arts; the Social type prefers working with people; the Enterprising type is interested in leading and persuading others; and the Conventional type prefers activities that involve dealing with structured data. Holland's theory assumes that these six interest dimensions are arranged in a hexagonal structure (see Figure 1). Within the hexagon, similar interests are located closely to each other, whereas conflicting interests take opposite positions. Holland proposed several hexagonal structural models all having the same underlying, fundamental hexagonal structure but more or less constrained in their assumptions. The most constrained model is the perfect circumplex model, where equal distances and consequently, equal correlations within adjacent interests, within alternate interests, and also within opposite interests are assumed. The higher the similarity between the interest scales measured by correlations, the shorter the distance between the interests scales (inversely proportional).

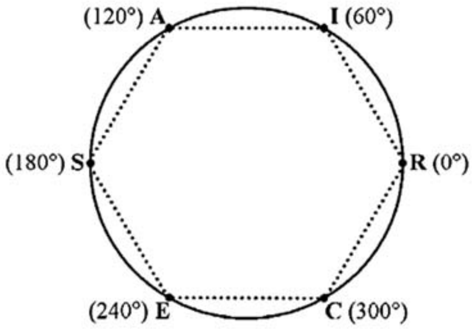

Figure 1. Graphical representation of Holland's hexagonal model (dotted lines) and its representation as a perfect circumplex model (solid lines). Reprinted from "The structure of vocational interests in Germany: Different methodologies, different conclusions" by G. Nagy, U. Trautwein and O. Lüdtke, 2010, *Journal of Vocational Behavior*, *76*(2), p. 154. Copyright 2007 by Elsevier. Reprinted with permission.

2.1.2 Hierarchical models by Gati and by Rounds and Tracey

In the research literature, various alternatives to Holland's circumplex model have been proposed. One of the best established competing models is the hierarchical model of vocational interests by Gati (1979, 1982, 1991). Gati's model relates to Holland's six personality types, and it is known as the three-group partitioning model because it separates the interest domains into three hierarchical groups: R-I, A-S, E-C (see Figure 2). Based on cluster analysis, its main assumption holds that dimensions within the three clusters are very similar. Regardless of any assumptions across clusters, individuals' interests within clusters correlate higher compared to non-paired domains (i.e., R-A, R-S, R-E, R-C, I-A, I-S, I-E, I-C, A-E, A-C, S-E, and S-C). As stated by Gati (1998), the hierarchical tree structure also reflects a top-down process of career-related decisions. Using a multistep classification, individuals first choose from few major clusters, and later, they continue with more refined distinctions based on the relatively less salient differences (Gati, 1991).

Gati (1991) argued that the Artistic domain should constitute a separate cluster. Later, as empirical evidence revealed that the correlation between the Artistic and the Social domain was low (e.g., Fouad & Dancer, 1992; J. L. Swanson, 1992; T. J. Tracey & Rounds, 1993), similar to Gati's model, Rounds and Tracey (1996) proposed a hierarchical partitioning model that separated the Artistic domain whereas the Social domain was included in the S-E-C cluster (see Figure 2). Again as in Gati's model, interest domains correlate higher within the clusters compared to across clusters (non-paired clustered domains).

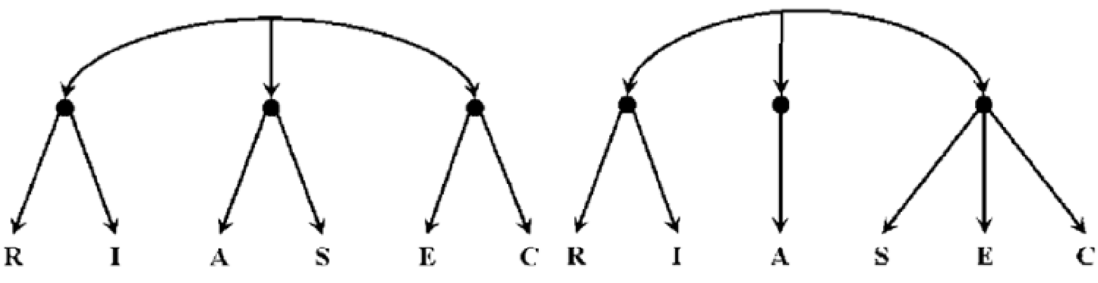

Figure 2. Graphical representation of Gati's hierarchical model (left panel) and Rounds and Tracey's hierarchical model (right panel). Arrows represent the assumed top-down process of career-related decisions. Reprinted from from "The structure of vocational interests in Germany: Different methodologies, different conclusions" by G. Nagy, U. Trautwein and O. Lüdtke, 2010, *Journal of Vocational Behavior*, *76*(2), p. 155. Copyright 2007 by Elsevier. Reprinted with permission.

2.1.3 Validity of vocational interest models

From a historical viewpoint, the two vocational interest models have undergone substantive validation. In responds to Holland's hexagonal model, Gati (1991) criticized Holland's model and stated that his model was superior. However, the structural meta-analysis by T. J. Tracey and Rounds (1993) demonstrated that Gati's comparison was flawed as incomplete specifications of Holland's model had been included which had led to incorrect order comparisons between both structural models. T. J. Tracey and Rounds' (1993) meta-analytical approach, in particular, supports Holland's hexagonal structure by indicating a good fit of the circumplex model that is also superior to Gati's hierarchical model. Research regarding the structure of vocational interests has its roots in the United States. In these studies, Holland's hexagonal model has been widely confirmed on different U.S. samples and for different target groups, as for example, no differences in the circular structure were found across racial-ethnic groups (African Americans, Mexican Americans, Asian Americans, Native Americans, and Caucasians) or college students versus 10th graders (Day, Rounds, & Swaney, 1998).

Towards the beginning of the 21st century, researchers were interested in validating Holland's interest structure in different cultural settings. Regarding generalizability across culture, some studies investigating Holland's structural assumptions outside the U.S. found evidence in support of Holland's hexagonal model (e.g., Darcy, 2005; Nagy, Trautwein, & Lüdtke, 2010). For example, in Germany, Nagy et al. (2010) reported that Holland's circumplex model was superior to hierarchical models (such as Gati's model and also Rounds and Tracey's model) tested. In line with Day et al. (1998), they found invariance between high school and university students. In constrast, other studies, especially meta-analyses, revealed that Holland's structural conceptualization was not supported in samples of cross-culture matrices (e.g., Long & Tracey, 2006; Rounds & Tracey, 1996).

Moreover, not only cultural comparisons revealed contradictory results, regarding the generalizability across gender, differences in the psychometric structure emerged also for female and male samples. Findings regarding gender differences are quite contrary: Some show evidence for invariance favoring Holland's hexagonal model across gender (e.g., Anderson, Tracey, & Rounds, 1997; T. J. Tracey & Rounds, 1993). In a large empirical study–nearly 70,000 U.S. high school students were investigated–Darcy and Tracey (2007) found that Holland's model fit both genders equally well. Further, Darcy and Tracey (2007) demonstrated that Holland’s assumptions are generalizable across age. However, other studies showed gender differences in the model structure (e.g., Armstrong, Hubert, & Rounds, 2003; Nagy et al., 2010). As, the gender issue is highly relevant to further investigations of the validity of vocational interests, research literature about gender differences are further described (section 2.3).

2.2 Vocational interests and cognitive abilities

Prior to focusing on the relations between vocational interests and cognitive abilities (section 2.2.1), cognitive abilities are defined, and an influential contemporary theory of cognitive abilities with its assumptions is described, that is, Cattell-Horn-Carroll (CHC) theory (Flanagan & Dixon, 2013; Schneider & McGrew, 2012). Afterwards, the research literature on the relationship between specifically Holland's RIASEC types and (a) general intelligence (section 2.2.2) and (b) specific cognitive abilities is reviewed (section 2.2.3).

The nature and structure of cognitive abilities were highly debated in the last century. In 1994, a group of experts in the field of cognitive ability research and related disciplines consented on the following definition of intelligence:

> "Intelligence is a very general mental capability that, among other things, involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience. It is not merely book learning, a narrow academic skill, or test taking smarts. Rather it reflects a broader and deeper capability for comprehending our surroundings–'catching on', 'making sense' of things, or 'figuring out' what to do" (L. S. Gottfredson, 1997, p. 13).

Furthermore, in the last decades, an understanding has emerged that cognitive abilities are organized hierarchically with a general factor, labeled *g* or intelligence or general mental ability, and a series of specific or primary cognitive abilities that are moderately correlated with the general factor (Carroll, 1993). Measures of general intelligence are effective predictors of job and academic performance (Kuncel, Hezlett, & Ones, 2004; Schmidt & Hunter, 2004). Specific abilities such as verbal, quantitative, and spatial abilities possess psychological importance beyond *g*, especially for predicting educational and vocational choices (Achter, Lubinski, Benbow, & Eftekhari-Sanjani, 1999; Humphreys, Lubinski, & Yao, 1993). The CHC theory of cognitive abilities is viewed as an influential contemporary theory of cognitive abilities (Flanagan & Dixon, 2013; Schneider & McGrew, 2012). By merging Horn-Cattell's extended *Gf*-*Gc* theory (Horn & Noll, 1997)–on fluid (*Gf*) and crystallized intelligence (*Gc*)–with Carroll's three-stratum theory (Carroll, 1993), CHC theory is effectively an amalgam of more than 60 years of factor-analytical research in the domain of cognitive abilities. In accordance with Carroll's three-stratum theory, CHC theory assumes a hierarchical model of cognitive abilities with three strata: General intelligence (*g*) is located at the apex (Stratum III), 16 broad cognitive abilities (e.g., fluid reasoning, visual processing, psychomotor abilities) are located at Stratum II, and more than 80 narrow abilities (e.g., perceptual speed, working memory capacity) at Stratum I (for detailed descriptions of broad and narrow abilities, see Schneider & McGrew, 2012).

According to Carroll (1993), broad abilities represent "characteristics of individuals that govern and influence a great variety of behaviors in a given domain," whereas narrow abilities represent "specializations of abilities … that reflect the effects of experience and learning, or the adoption of particular strategies to perform" (p. 634). Recent studies highlight the invariance of CHC broad ability factors across different intelligence batteries (Reynolds, Vannest, & Fletcher-Janzen, 2013) and its usefulness as a framework for classifying intelligence and achievement batteries (Flanagan, Alfonso, & Ortiz, 2012). Moreover, the CHC model is perceived as the most empirically supported and theoretically sound model of the structure of human cognitive abilities (Alfonso, Flanagan, & Radwan, 2005; Stankov, 2000), thus emphasizing the CHC model's advantage as a classification system for meta-analyses.

2.2.1 Relations between interests and cognitive abilities

Interests are perceived as antecedents of performance. In his investment theory, Cattell (1971, 1987) assumed that individuals differ in their initial level of fluid intelligence that is genetically and neurophysiologically based. Hence, fluid intelligence is the main determinant of performance in infancy. Across the life span, individuals invest their fluid intelligence in the development of skills and acquisition of knowledge (i.e., crystallized intelligence). According to Cattell, this development is not only driven by availability and quality of education, family resources, effort, motivation, and ambition but also by an individual’s interests. Or as Strong (1943) suggested, “the relationship among abilities, interests, and achievement may be linked to a motor boat with a motor and a rudder. The motor (abilities) determines how fast the boat can go, the rudder (interests) determine which way the boat goes” (p. 17). Thus, whereas cognitive abilities predict performance because they determine what individuals “can do,” interests direct where one’s intellectual potential is invested. In line with Cattell (1971, 1987), Ackerman (1996) proposed in his process, personality, interests, and knowledge theory that the development of intelligence (i.e., the transition from intelligence-as-process to intelligence-as-knowledge) is guided by motivation, personality, and also interests. Further, intelligence-as-knowledge is thought to form the core of adult intelligence.

J. L. Holland (1997) assumed that children select those activities that match their abilities and their preferences which are either genetically determined or chosen by their parents. As these activities constantly receive positive reinforcement, they lead to an increasing stabilization of individual preferences as well as to an integration of interests, abilities, and skills during adolescence and early adulthood. The idea that people are interested in things they are good at and vice versa seems convincing; however, empirical studies demonstrated only weak to moderate correlations between interests and abilities (e.g., Ackerman & Heggestad, 1997; Carless, 1999; Proyer, 2006; Randahl, 1991).

In his integrative theoretical model for individual differences, Schmidt (2014) highlighted that both general interests, such as typical intellectual engagement (TIE), and specific interests, such as Holland’s occupational interests, should predict academic and occupational performance by guiding the development of crystallized intelligence (i.e., general and specific knowledge and skills). Recent empirical findings support these assumptions. In their meta-analysis, both Van Iddekinge et al. (2011) and Nye et al. (2012) demonstrated that specific interests predicted academic and occupational performance. Furthermore, recent research showed that various investment traits such as TIE or need for cognition positively correlate with crystallized intelligence, academic performance, and acquired knowledge (Von Stumm & Ackerman, 2013; Von Stumm, Hell & Chamorro-Premuzic, 2011). However, the relation between what Schmidt (2014) defined as specific interests and both fluid and crystallized intelligence should be investigated more closely.

Despite some efforts to consider the relationship between intelligence and vocational interests (e.g., Carson, 1998a; Lowman & Leeman, 1988; Proyer, 2006; Randahl, 1991), most studies addressed the issue from the perspective of self-estimated abilities. However, self-ratings of abilities are susceptible to self-presentational biases and are only moderately correlated with objectively assessed or measured cognitive abilities (Zell & Krizan, 2014). Furthermore, the reported overlap between interests and self-estimated abilities may partly be explained by common-method variance (Lowman & Carson, 2013). In their meta-analysis, Ackerman and Heggestad (1997) focused on the overlap between cognitive abilities, vocational interests, and personality. Due to the small number of studies that reported correlations between vocational interests and cognitive ability measures, Ackerman and Heggestad had to rely on a qualitative review. Summarizing patterns in interest-ability correlations from five studies, they concluded that there are only moderate correlations between specific cognitive abilities and vocational interests.

2.2.2 Holland's RIASEC types and general cognitive ability

Although J. L. Holland (1973) proposed that different occupational types have developed a characteristic repertoire of skills, competencies, and abilities, there are relatively few references on the precise relationship between vocational interests and cognitive abilities in his work. However, some additional indications can be drawn from J. L. Holland's earlier work (1959) where the six occupational types were still labeled motoric, intellectual, esthetic, supportive, persuasive, and conforming. For example, J. L. Holland (1959) described persons with an intellectual orientation as "task-oriented people who generally prefer to 'think through', rather than to 'act out', problems. They have marked needs to organize and understand the world" (p. 36). Later J. L. Holland (1973, 1985) referred to the Investigative type as scholarly and intellectual and proposed that the Investigative type has higher levels of general intelligence than the Realistic and Artistic type. Contrary, persons with a supportive orientation are assumed to "avoid situations requiring intellectual problem solving" (J. L. Holland, 1959, p. 37).

In his interdomain career assessment model, Lowman (1991) made a first attempt to systematically review the relationship between interest themes and cognitive abilities. Lowman related high levels of intelligence with the Investigative type, moderate levels of intelligence with the Social and Enterprising types, and low to average intelligence levels with the Realistic and Conventional types. Empirically, there is strong evidence for a positive relation between Investigative interests and *g* (Carson, 1998a; Proyer, 2006; Reeve & Heggestad, 2004). Additionally, analyses of occupational data showed that investigative occupations require the highest level of *g* (L. S. Gottfredson, 1986). Empirical evidence further points to a positive correlation between Artistic interests and *g* (Carson, 1998a; Proyer, 2006; Reeve & Heggestad, 2004). Unfortunately, evidence from past research is less definite regarding the relation between *g* and Realistic, Social, Enterprising, and Conventional interests.

2.2.3 Holland’s RIASEC types and specific cognitive abilities

J. L. Holland (1959, 1973) proposed that each RIASEC type is characterized not only by specific interests but also by specific abilities and competencies.

Holland assumed persons “activities requiring physical strength, aggressive action, motor coordination and skill” (J. L. Holland, 1959, p. 36), and further related the Realistic type with mechanical abilities and a lack of social skills. According to Lowman’s (1991) review, Realistic interests should further be positively related to spatial abilities, and negatively related to verbal abilities. Furthermore, Ackerman and Heggestad (1997) concluded that positive relations tend to be found between Realistic interests and spatial, mathematical, and mechanical abilities. Recent research supports the positive correlation between Realistic interests and spatial abilities (Carson, 1998b; Proyer, 2006) as well as mathematical abilities (Carson, 1998b).

Investigative interests were found to have positive correlations with spatial, mathematical, and also verbal abilities (Ackerman & Heggestad, 1997); assumptions supported by recent research (Carson, 1998b; Proyer, 2006). Furthermore, Lowman (1991) proposed that Investigative interests are associated with high levels of reasoning and convergent thinking. J. L. Holland (1959, 1973) associated the Investigative type with mathematical and scientific abilities but also a lack of leadership abilities.

J. L. Holland (1959) assumed that persons with an esthetic orientation “prefer dealing with environmental problems through self-expression in artistic media” and highlighted that they “avoid problems requiring interpersonal interactions, a high degree of structuring, or physical skills” (p. 37). Later J. L. Holland (1973) associated the Artistic type with verbal abilities as well as divergent thinking. Lowman (1991) further proposed a positive relationship between Artistic interests and spatial abilities. Neither of these assumptions was supported by recent research (Carson, 1998b; Proyer, 2006).

J. L. Holland (1959) assumed that persons with a supportive orientation have verbal and interpersonal skills and “avoid situations requiring intellectual problem-solving, physical skills or highly ordered activities” (p. 37). J. L. Holland (1973) further related the Social type with a lack of mechanical and scientific abilities. Empirically, Social interests were found to be uncorrelated or negatively correlated with specific cognitive abilities (Ackerman & Heggestad, 1997). Ackerman (1997) proposed that this could be seen as evidence that cognitive ability measures insufficiently capture domains such as social or interpersonal abilities. Recent research reported negative relations with verbal, numerical, and spatial abilities (Carson, 1998b; Proyer, 2006).

J. L. Holland (1959) indicated that persons with a persuasive orientation “prefer to use their verbal skills…for dominating, selling, or leading others”. Lowman (1991) assumed positive relations with interpersonal as well as management abilities. Ackerman and Heggestad (1997) concluded in their review that negative associations tend to be found between ability measures and Enterprising interests. This assumption is further supported by recent research (Carson, 1998b; Proyer, 2006).

The Conventional type is associated with clerical and numerical abilities (J. L. Holland, 1973) as well as with computational abilities and perceptual speed (Lowman, 1991), but avoids “ambiguous situations or problems involving interpersonal relationships and physical skills” (J. L. Holland, 1959, p. 37). Ackerman and Heggestad (1997) as well as Carson (1998a) confirmed a positive relation between Conventional interests and mathematical computation as well as perceptual speed.

2.3 Vocational interests and gender differences

Considerable mean differences are consistently found in vocational interests of women and men (e.g., Lippa, 1998; Su, Rounds, & Armstrong, 2009). On average, whereas women prefer working with people, men are drawn to thing-oriented activities. Speaking within the Realistic, Investigative, Artistic, Social, Enterprising, and Conventional (RIASEC) framework (J. L. Holland, 1997), men show stronger interest in the Realistic and Investigative domain, while women express stronger interest in the Artistic, Social, and Conventional domain. In their meta-analysis, Su et al. (2009) reported an effect size of d = 0.93 for the people-things dimension—one of the largest sex difference found in the psychological domain (Hyde, 2005).

Gender differences in cognitive abilities and the question of how gender influences the relationship between vocational interests and cognitive abilities are described next (section 2.3.1). Then (section 2.3.2), the question of where these large gender-related mean differences within the vocational interest domain originate from leads to the issue of measurement bias. These gender differences in vocational interests might be interpreted as an indicator of unfairness in interest inventories. Approaches of test fairness are described.

2.3.1 Gender differences in the relationship between interests and cognitive abilities

The extent to which the relationship between vocational interests and cognitive abilities is moderated by gender is debated controversially. As gender differences in vocational interests were presented previously, gender differences in cognitive abilities are presented next. Then, the interrelation between vocational interests and cognitive abilities is described.

Although research showed negligible gender differences on general intelligence (Deary, 2003; Strand, Deary, & Smith, 2006; Halpern, 2000), males tend to perform better in some subtests and females on others. Meta-analyses on verbal abilities (Hedges & Nowell, 1995; Hyde & Linn, 1988) found small to moderate differences favoring females on reading comprehension, writing, and speech production but not all tests of verbal ability. Likewise, meta-analyses on mathematical abilities (Hedges & Nowell, 1995; Lindberg, Hyde, Petersen, & Linn, 2010) found negligible to small overall differences with females performing better at measures of mathematical computation and males performing better at mathematical problem solving tasks. Moderate to large differences favoring males were found on measures of mental rotation, spatial perception, and mechanical reasoning (Lemos, Abad, Almeida, & Colom, 2013; Schmidt, 2011; Voyer, Voyer, & Bryden, 1995). A study investigating over 1.6 million 7th grade students in the right tail (top 5%) in ability across 30 years found that the ability ratio in mathematical reasoning favors males, and it showed stability over the last 20 years even though it is substantially lower now than 30 years ago. Results in favor of females revealed higher scores in verbal reasoning and writing ability tests (Wai, Cacchio, Putallaz, & Makel, 2010). Furthermore, research suggested that (a) gender differences in cognitive abilities vary by age (Lindberg et al., 2010), and (b) males show greater variability than females on most cognitive ability measures (Deary, 2003; Hedges & Nowell, 1995; Johnson, Carothers, & Deary, 2008; Strand et al., 2006).

Further, gender differences in vocational interests are perceived as antecedents of gender differences in the development of skills, knowledge, and aptitudes: Schmidt (2011), for example, proposed that gender differences in technical interests lead to differences in technical experiences and technical knowledge acquisition, which in turn lead to gender differences in technical aptitude. Similarity of verbal and quantitative aptitude, in contrast, results from common experiences during formal education. Although few studies (Carless, 1999; Reeve & Heggestad, 2004) examined the relationship between interests and cognitive abilities by gender, results suggested that the direction and magnitude of correlation coefficients differ to some extent for females and males. Further, Johnson and Bouchard (2009) found that gender differences in cognitive abilities partially explained differences in vocational interests. In their review on intelligence, interests, and personality, Ackerman and Heggestad (1997) noted that not considering gender differences is a specific limitation.

2.3.2 Gender differences in the measurement of vocational interests

Facing one of the largest gender differences found in the psychological domain (Hyde, 2005), the search of a possible origin of these gender differences has become a widely discussed question. Therefore, it is debated why women and men differ in their vocational interests: Social theories emphasize the influence of gender socialization and gender roles (L. S. Gottfredson, 1981, 2004; Ruble, Martin, & Berenbaum, 2006), whereas biological approaches point to the importance of genes as well as pre- and postnatal hormones and their influence on the brain structure and neuronal development (Berenbaum, Baxter, Seidenberg, & Hermann, 1997; Cohen-Bendahan, Van de Beek, & Berenbaum, 2005; Goy & McEwen, 2007). Current research on gender differences in diverse cultures (Lippa, 2010) and on the impact of prenatal and postnatal hormones on gender-specific interest pattern (Hell & Pässler, 2011) questions the assumption that the vocational interests of women and men are fundamentally equal and that empirically reported mean differences merely rest upon measurement errors or solely reflect differences in socialization history.

In their meta-analysis, Su et al. (2009) showed that item development (i.e., eliminating items with large gender-specific mean differences) moderated the size of sex differences in Realistic, Investigative, and Enterprising interests. In comparison, the reduction in gender differences was minimal for the Artistic, Social, and Conventional domains, thus indicating that efforts to reduce gender differences in vocational interest inventories have been primarily successful with scales traditionally favoring males. In line with the opportunity approach of validation (Prediger & Cole, 1975), interest inventories applying this item development technique aim at encouraging women to explore nontraditional educational and occupational choices such as careers in engineering and science. Nevertheless, the question remains unsettled whether eliminating items with large gender differences might result in changes of the construct the interest inventory is intended to measure, and thereby reducing its predictive validity (G. D. Gottfredson & Holland, 1978).

Su et al. (2009) emphasize that the attempt to remove gender differences from interest inventories results in scales that do not mirror the original RIASEC dimensions as proposed by J. L. Holland (1959, 1997), but represent a narrower range of interests. Russell (2007) examined the agreement between four interest inventories and showed a mere 50% cross-classifying hit rate of RIASEC codes between the Self-Directed Search (SDS; J. L. Holland, Fritzsche, & Powell, 1994) and the Revised Unisex Edition of the American College Testing (UNIACT-R; American College Testing [ACT] Program, 1995). Thus, depending on the interest inventory applied and its construction strategy (i.e., whether items with gender-specific mean differences are eliminated), a person would likely receive different occupational suggestions. That indicates that item development strategies might indeed influence instrument validity.

One question highly debated in the field of vocational interests is whether structural presumptions hold for both females and males. Past studies yielded contradictory results: Some show evidence for invariance across gender (e.g., Anderson et al., 1997; Darcy & Tracey, 2007; Nagy et al., 2010; T. J. Tracey & Rounds, 1993), other studies reveal considerable gender differences in model structure (e.g., Armstrong et al., 2003; Hansen, Collins, Swanson, & Fouad, 1993). Thus, the issue of structural gender invariance remains an unresolved matter. Moreover, up till now it remains unclear how structural properties of the interest inventory are impacted by those test development strategies (i.e., eliminating items showing large gender-specific DIF). Structural validity, however, is an important facet of construct validity.

However, both the Guidelines for the Assessment for Sex Bias and Sex Fairness in Career Interest Inventories (National Institute of Education [NIE], 1975) and the Standards for Educational and Psychological Testing (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education [AERA, APA, & NCME], 1999), strongly emphasize the suspension of gender differences in instrument validity and prediction as one crucial aspect of test fairness. According to the Standards, “no bias exists if the regression relating the test and the criterion are indistinguishable for the groups in question” (AERA, APA, & NCME, 1999, p. 79). In recent decades, the Cleary (1968) approach of testing for differences in regression lines has been the central approach to evaluate test bias in prediction.

Moreover, evaluating the Cleary approach, Meade and Tonidandel (2010) demand that when testing for bias, both internal and external approaches must be considered. Therefore, first measurement biases must be ruled out by testing for differential item and test functioning. Second, by testing for differential validity and differential prediction, the absence of bias in prediction must be demonstrated. While differential validity determines whether the correlation between a predictor and a criterion is equal for women and men, differential prediction refers to differences in the standard errors of estimates or regression lines (i.e., differences in the slopes and/or intercepts) between females and males (Linn, 1978). Importantly, though differential validity and differential prediction are related, they can occur independently of each other. Thus, a test may predict a criterion with the same accuracy for females and males, but the use of a single regression line for females and males would lead to overprediction or underprediction for one group. Thus, as Linn (1978) argues, differential prediction is the more crucial aspect of instrument validity because it has a more direct influence on considerations of fairness.

Internal Approach of Test Fairness: DIF in the Interest Domain

Item response theory (IRT) has provided new methods to address the issue of fairness in interest inventories. Within the framework of IRT, it is possible to determine whether women and men respond differently to the same item, given that they possess the same underlying trait level (P. W. Holland & Wainer, 1993; Osterlind & Everson, 2009). Multidimensional IRT has been developed as a theoretical framework to examine how DIF relates to item and test validity. Thus, the target trait, or primary dimension, is distinguished from other factors influencing test performances, that is, nuisance factors and secondary dimensions. DIF occurs because a factor apart from the construct targeted by the test is affecting the endorsement of one group (e.g., women) but not of the other group (e.g., men). Consequently, gender-specific DIF in interest items indicates that women and men do not show the same probability of endorsing an item (e.g., *teach children*) despite having the same underlying interest level (e.g., Social interest).

Testing for DIF is relatively rare in the domain of vocational interests, even though it is a standard procedure in ability testing. Current research (Aros, Henly, & Curtis, 1998; Einarsdóttir & Rounds, 2009) on gender differences in interest tries to establish whether these gender-related differences are attributable to real and valid differences in the underlying trait or whether they additionally depend upon certain properties of the instrument being administered. Aros et al. (1998) found gender-specific DIF on most of the Strong Interest Inventory (SII; Harmon, Hansen, Borgen, & Hammer, 1994) occupational title items, whereas Einarsdóttir and Rounds (2009) detected gender-specific DIF in two thirds of the SII items, especially on the Realistic scale. Both Aros et al. (1998) and Einarsdóttir and Rounds (2009) argued that responses on the SII were influenced by an additional sex-type dimension leading to DIF. When biased items were removed, gender mean differences favoring men were reduced in the Realistic and Investigative scales, whereas gender differences favoring women remained on the Artistic and Social scales. Comparing different test construction strategies, Wetzel, Hell, and Pässler (2012) concluded that IRT-based DIF analyses are useful for test constructing purposes.

External Approach of Test Fairness: Differential Validity and Differential Prediction in the Interest Domain

According to J. L. Holland (1997), individuals tend to gravitate to environments that are consistent with their interest profile. For example, individuals working in sciences (e.g., biology, physics, and chemistry) are expected to have predominantly Investigative interests. Therefore, instrument validity is evaluated by comparing the predominant interest type (high-point code) with a criterion, such as educational or occupational choice, and calculating the percentage of agreement. Various studies analyzed the concurrent or predictive validity of interest inventories by examining hit rate differences for women and men. Most studies found comparable mean hit rates for females and males (ACT, 2009; Hansen & Swanson, 1983; Hansen & Tan, 1992; Hansen & Dik, 2005; Harrington, 2006). Other researchers report slightly better classification results for women (G. D. Gottfredson & Holland, 1975). Furthermore, although mean hit rates for women and men were shown to be comparable, larger differences are found when comparing gender-specific hit rates on occupational group level (ACT, 2009). Another procedure to establish the level of agreement between a person and the environment is Holland’s code comparisons based on congruence indices. The C-Index (Brown & Gore, 1994) has been frequently recommended as congruence index because it not only incorporates Holland’s circular order assumption but is also sensitive to the order of interest types in Holland codes assigned for persons and occupations (Dik, Hu, & Hansen, 2007). High level of person-environment congruence will increase the likelihood of continuity in educational and occupational decisions as well as the satisfaction and success in this type of educational or occupational environment (J. L. Holland, 1997). Applying this procedure, Rottinghaus, Coon, Gaffey, and Zytowski (2007) found no indication for gender differences in the level of congruence. Remarkably, whereas there is extensive literature investigating gender-related differential validity and differential prediction in the cognitive ability domain as well as studies on differential validity in the interest domain (ACT, 2009; Hansen & Swanson, 1983; Hansen & Tan, 1992; Hansen & Dik, 2005; Harrington, 2006; Rottinghaus et al., 2007), the issue of differential prediction has so far been left unexamined.

Since important educational and occupational decisions are also based upon results of interest tests, studying this form of predictive bias in vocational interest inventories seems particularly important.

CHAPTER 3 EMPIRICAL EVIDENCE FROM STUDIES ON VOCATIONAL INTERESTS

In order to strive for career direction in the exploration stage, individuals have to make vocational choices. Such vocational choices are guided by vocational interests in this early career stage, which provide the basis for individuals' career construction in the following career stages. Recent meta-analyses provide critical evidence for the predictive validity of interests for performance criteria in both work and academic settings (e.g., Nye et al., 2012; Van Iddekinge et al., 2011). With regard to this, it seems important to investigate current research questions in the field of vocational interests.

First specific research gap: relationship between vocational interests and both general as well as specific cognitive abilities (Study 1)

Many studies on vocational interest research investigated how vocational interests relate to established predictors of job performance (i.e., cognitive abilities and personality). The relationship between vocational interests and personality received considerable attention in comprehensive meta-anlyses (e.g., Barrick, Mount, & Gupta, 2003; Larson, Rottinghaus, & Borgen, 2002; Staggs, Larson, & Borgen, 2007), whereas no comprehensive meta-analytic review has been conducted to analyze the question of how vocational interests and both general as well as specific cognitive abilities are related. As vocational interests and cognitive abilities have been identified as key factors for choosing a career, it seems interesting to explore the relationship between both constructs. As the first study provides answers to this question, it contributes to this specific research gap. The first study meta-analytically investigates the nature and magnitude of the relationship between vocational interests and both general as well as specific cognitive abilities. Furthermore, it is explored whether this relationship varies by gender, age, and birth cohort as possible moderators (section 3.1).

Second specific research gap: comparison of differential validity and differential prediction between a standard and a DIF-optimized vocational interest inventory (Study 2)

Since important educational and vocational choices can be based on results of interest inventories, it is immensely important to adequately measure vocational interests with the lowest measurement bias possible. In response to NIE and APA Standards (NIE, 1975) that strongly emphasize the elimination of gender differences in instrument validity and prediction, the question arises whether eliminating interest items showing large DIF can reduce gender differences on the scale-level. How can test fairness be improved in vocational interest inventories? As the second study provides answers to these questions, it contributes to this specific research gap. The second study addresses the issue of gender-related differential validity and differential prediction in vocational interest inventories as one major concern of test fairness (section 3.2).

Third specific research gap: effects of gender-specific DIF on the psychometric structure of a standard RIASEC interest inventory (Study 3)

Further to the overarching goal of ensuring test fairness of vocational interest inventories across genders, measures to correct for the measurement bias (e.g., DIF analyses) need to ensure that theoretical structural properties remain invariant, for example, when eliminating items showing large DIF. This leads to the question: How does the elimination of items with DIF for males and females affect the structural validity of vocational interest measures? As the third study provides answers to this question, it contributes to this specific research gap. Addressing this question not only informs research on interest structure but also has great practical implications for interest assessment and the development of gender-balanced and fair interest measures, thus contributing to the literature of interest measurement and gender fairness. The third study examines how gender-specific DIF affects the psychometric structure of a standard RIASEC interest inventory (section 3.3).

Chapter outline

Contributing to the literature of career choice in early career stages, current research questions or issues in the field of vocational interests as mentioned above are addressed in chapter 3. Each of the three studies on vocational interests is separately presented according to APA standard research format: aims and hypotheses of the study; methodology including a description of participants, instruments, and procedures; results; and discussion. Within the discussion, study results are summarized and embedded within the research literature. Limitations of each study and implications for future research and practice are discussed.

3.1 Vocational interests and intelligence: a meta-analysis

3.1.1 Aims and hypotheses

Aims

Vocational interests are established predictors of educational choice (Hansen & Neuman, 1999), degree completion (Webb, Lubinski, & Benbow, 2002), occupational choice (Hansen & Dik, 2005), and occupational satisfaction (Tsabari, Tziner, & Meir, 2005). Their importance in personnel selection and their relevance for understanding performance has often been questioned in past research (Barrick & Mount, 2005; Hunter & Hunter, 1984). However, recent meta-analyses (Nye et al., 2012; Van Iddekinge et al., 2011) called for a reconsideration of interests for predicting performance-relevant criteria. Van Iddekinge et al. (2011) demonstrated the importance of interests for predicting job and training performance as well as turnover. Nye et al. (2012) showed that interests are related to performance and tenure not only in work but also in academic contexts. In addition, prediction of performance was strongest when academic or work environment matched individuals' interests. Thus, these meta-analyses provide critical evidence for the predictive validity of interests for performance criteria in both work and academic settings.

This renewed attention to vocational interests also raises the question of how vocational interests relate to established predictors of job performance (i.e., cognitive abilities and personality). Whereas the relation between interests and personality received considerable attention in both personnel selection and vocational choice literature (Barrick et al., 2003; Larson et al., 2002; Staggs et al., 2007), no comprehensive quantitative summary has thus far been conducted to analyze the relationship between vocational interests and cognitive abilities. Therefore, the main purpose of this study is to address this gap and systematically examine the nature and magnitude of the relation between these two constructs.

The study is based on the following publication:
Pässler, K., Beinicke, A., & Hell, B. (2015). Interests and intelligence: A meta-analysis. *Intelligence*, *50*, 30-51. doi:10.1016/j.intell.2015.02.001

Hypotheses

Given its wide proliferation and profound empirical support, we referred to Holland's RIASEC framework to examine vocational interests. As reviewed in closer detail above (section 2.2), the expected relations between Holland's occupational themes and cognitive abilities lead to the following hypotheses:

Holland's RIASEC types and general intelligence (g)

In line with J. L. Holland's (1959, 1973, 1985) assumptions, we proposed that:

Hypothesis 1: *Investigative interests will be positively related to g (Hypothesis 1a) and that the relation between Investigative interests and g will be stronger than the relationship with any other interest type (Hypothesis 1b).*

Holland's RIASEC types and specific cognitive abilities

Based on J. L. Holland's (1959, 1973) assumptions, we hypothesized that:

Hypothesis 2: *Realistic interests are positively related to mechanical abilities (Hypothesis 2a) as well as motor coordination (Hypothesis 2b).*

J. L. Holland (1959, 1973) associated the Investigative type with mathematical and scientific abilities but also a lack of leadership abilities. Therefore, we hypothesized that:

Hypothesis 3: *Investigative interests are positively related to numerical abilities (Hypothesis 3a) and induction (Hypothesis 3b).*

Again, following Holland's assumptions, we hypothesized that:

Hypothesis 4: *Artistic interests are positively related to verbal abilities (Hypothesis 4).*

Based on Holland’s assumptions, we hypothesized that:

Hypothesis 5: *Social interests will be positively related to verbal abilities (Hypothesis 5a) but negatively related to mechanical abilities (Hypothesis 5b).*

In line with Holland, we hypothesized that:

Hypothesis 6: *Enterprising interests will be positively related to verbal abilities (Hypothesis 6).*

Ackerman and Heggestad (1997) as well as Carson (1998a) confirmed a positive relation between Conventional interests and mathematical computation as well as perceptual speed. We hypothesize that:

Hypothesis 7: *Conventional interests will be positively related to numerical abilities (Hypothesis 7a) as well as perceptual speed (Hypothesis 7b).*

Moderators in the relationship between interests and cognitive abilities

Gender differences in interests and cognitive abilities

In light of presented findings as reviewed in detail above (section 2.3.1) and the fact that Ackerman and Heggestad (1997) noted not considering gender differences as a specific limitation of their review on intelligence, interests, and personality, we investigated gender as a possible moderator in the relationship between vocational interests and cognitive abilities.

Age differences in interests and cognitive abilities

In their meta-analysis of longitudinal studies on rank-order and profile stability, Low, Yoon, Roberts, and Rounds (2005) demonstrated that vocational interests are relatively stable, even at early adolescence. Moreover, interest stability greatly increases in early adulthood and then remains stable for the next two decades. This marked stability increase in early adulthood is assumed to result from fewer environmental constraints since individuals at this age typically leave their familiar surrounding for novel settings such as college or work places (Low & Rounds, 2007), thus enabling individuals to choose environments and activities that match their vocational interests.

Although general intelligence (*g*) remains stable over time (Deary, Pattie, & Starr, 2013), there are some cognitive abilities that are more stable than others. Fluid intelligence has been found to increase throughout young adulthood, peaking in middle adulthood, and afterwards declining steadily (Kaufman, Johnson, & Xin Liu, 2008). This decline in fluid intelligence has been attributed to declines in processing speed and working memory (Kaufman et al., 2008). In contrast, crystallized ability has been found to increase with age throughout adulthood (Schaie, 2013). As highlighted previously, Cattell's investment theory (1971, 1987) assumes an age-related differentiation of cognitive abilities such that fluid intelligence is invested in the elaboration and formation of crystallized intelligence. Environmental and non-cognitive variables (e.g., motivation and interests) guide this knowledge acquisition. Therefore, we assumed that vocational interests and measures of crystallized intelligence should become closer related as individuals grow older, and the relation between vocational interests and those cognitive abilities that are highly influenced by experience and knowledge acquisition in the course of parental upbringing and education would be more pronounced in older samples than in younger samples. Within the CHC theory framework, domain-specific knowledge (e.g., mechanical knowledge or foreign language proficiency), quantitative knowledge, reading and writing, as well as language development are perceived as acquired knowledge constructs (Schneider & McGrew, 2012). Thus, we investigated individuals' age as a possible moderator in the relationship between vocational interests and cognitive abilities.

Cohort differences in interests and cognitive abilities

Recent research further indicated that vocational interests are influenced by cohort effects (Bubany & Hansen, 2011; Su et al., 2009). Investigating change across birth cohorts of college students, Bubany and Hansen (2011) found that although Enterprising and Social interests increased from earlier generations to more recent generations for both females and males, the increase in Enterprising interests was especially great in females. Further, for males, Bubany and Hansen (2011) reported a decrease in Realistic, Investigative, and Artistic interests. Moreover, gender differences in Investigative, Enterprising, and Conventional interests decreased from earlier generations to more recent generations. Similarly, Su et al. (2009) revealed that gender differences in Artistic and Enterprising interests were smaller for younger generations than older generations. These birth cohort changes are assumed to result from changes in the labor market, especially a steady increase in the number of women entering the workforce as well as an increase in college and graduate degrees earned by women (Bubany & Hansen, 2011). Thus, we considered birth cohort changes a potential moderator between interests and cognitive abilities, especially when relations are investigated separately for females and males.

3.1.2 Methodology

Literature search

To identify relevant (published or unpublished) literature for this meta-analysis, we searched the following databases: PsycINFO, PSYNDEX, ERIC, Academic Search Premier, and Business Source Premier. We searched titles, abstracts, or keywords of articles using combined keywords including the following terms and Boolean operators: (vocational preference OR vocational preferences OR vocational interest OR vocational interests OR occupational interest OR occupational interests OR occupational preference OR occupational preferences OR Holland* OR RIASEC* OR hexagon*) AND (cognitive ability OR cognitive abilities OR general mental ability OR general mental abilities OR aptitude* OR intelligence* OR ability*). Since the 1970s, interest literature has primarily used Holland's RIASEC taxonomy to organize research results on vocational interests (Armstrong, Su, & Rounds, 2011). Thus, only articles published after 1970 were investigated. We further reviewed the reference sections of those articles obtained by database search to identify additional articles. Finally, we contacted authors in the research field of vocational interests for unpublished data or work in progress.

Inclusion criteria

All primary studies were reviewed for meeting the following inclusion criteria: (a) a vocational interest inventory using Holland's RIASEC taxonomy, (b) cognitive ability measures based on objectively assessed (not self-reported) data, and (c) sufficient data (e.g., sample size, correlation coefficients) provided to compute effect sizes. If possible, we contacted the authors to obtain missing information. As Ackerman and Beier (2003) highlighted, vocational interest measures traditionally generate either similarity indexes or dominant typological themes. Thus, measures seldom yield continuous scores for individuals, making it impossible to compute correlations between vocational interests and cognitive abilities. Relatively few studies examine the association between objective cognitive abilities and vocational interests; instead they rely on self-estimated abilities. Overall, 27 studies representing 29 independent samples met all criteria of inclusion (see Table 1). All but two studies were published in peer-reviewed journals.

Table 1. *Overview of the Meta-analysis Database*

ID	Author(s)[a]/Article	Country	Sample size			Population	Moderator values			Vocational interest measure	Cognitive ability measure
			N	Male	Female		Age	Sex	Cohort		
1	Ackerman (2000)	United States	228	78	150	2	0	1	1960s	UNIACT	Ability battery
2	Ackerman et al. (2001)	United States	320			2	1	0	1980s	UNIACT	Ability battery
3	Ackerman et al. (1995)	United States	93	42	51	2	2	1	1970s	UNIACT	Ability battery
4	Bergmann (2013)*	Austria	5,134	2,269	2,866	1	1	1	1980s	GIST-R	KFT 4-12+R
5	Carless (1999a)	Australia		669	206	4	2	1	1960s	SDS	PL-PQ
6	Carless (1999b)	Australia		48	91	4	0	1	1970s	SDS	WAIS-R
7	Carson (1996)	United States	117			4	0	0	1970s	SII	DAT-A
8	Carson (1998a)	United States	547			4	0	0	1950s	SII	BAB
9	Carson (1998b)	United States	198			1	1	0	1970s	SDS	BAB
10	Fritzsche et al. (1999)	United States	90			4	0	0	1970s	SDS	WPT
11	Kanfer & Ackerman (1996)	United States	158			2	0	0		UNIACT	Ability battery
12	Kaub et al. (2012)	Germany	219	71	148	2	2	1	1980s	GIST-R	LPS-K
13	Kelso et al. (1977)	United States			192	1	0	1	1960s	SDS	ASVAB
14	Kirchler (1990)	Germany	86			4	0	0		GIST	BIST
15	Krapic et al. (2008)	Croatia	132			4	0	0	1970s	SDS**	APM
16	Lowman et al. (1985)	United States			149	2	0	1	1960s	SDS	Ability battery
17	Marcus et al. (2009)	Germany	268			3	1	0	1990s	GIST	WPT
18	Mussel (2013)	Germany	250	92	158	5	0	1	1980s	VPI	S&F
19	Pässler & Hell (2012)	Germany	1,990	809	1,181	2	2	1	1980s	WSI	Ability battery
20	Proyer (2006)	Austria	138	39	99	2	2	1	1970s	GIST	ISA
21	Randahl (1991)	United States	846			4	0	0	1940s	SVIB-SCII	GATB
22	Reeve & Heggestad (2004)	United States		16,010	20,443	4	0	1	1940s	VPI, SDS	Ability battery
23	Rolfhus & Ackerman (1996)	United States	180			2	1	0	1970s	UNIACT	Ability battery
24	Schmidt et al. (1998)	United States	695			1	0	0	1980s	SVIB-SCII	Ability battery
25	Stanley et al. (1995)	United States		188	90	1	0	1	1960s	HOC	DAT
26	Toker & Ackerman (2012a)	United States	184	82	102	2	0	1	1990s	UNIACT	ETS KIT

27 Toker & Ackerman (2012b)	United States	240	123	117	2	0	1	1990s	UNIACT	ETS KIT
28 Van Iddeking et al. (2011)	United States	418			4	2	0	1980s	WPS	AFQT
29 Vock et al. (2013)	Germany	4,680	2,123	2,557	1	1	1	1980s	GIST	KFT 4-12+R

Note. [a] Complete references can be found in the reference section. * unpublished data. ** Croatian version of the SDS. Sample sizes are presented for (total, male, female) samples that are included in the analyses.

In the coding of the population, 1 represents high school samples, 2 represents college or university samples, 3 represents apprentices, 4 represents workers, and 5 represents mixed samples. In the coding for age as a moderator, 1 represents a mean sample age smaller than 20 years with standard deviation less or equal than 5 years, 2 refers to a mean sample age greater or equal than 20 years with standard deviation less or equal than 5, and 0 represents either data with standard deviations greater than 5 or no available data. For sex as a moderator, female- or male-specific samples were coded as 1, whereas data with only the total sample available was coded as 0.

Interest measures: GIST = General Interest Structure Test, GIST-R = General Interest Structure Test-Revised, HOC = Holland Occupations Checklist, SDS = Self-Directed Search, SII = Strong Interest Inventory, SVIB-SCII = Strong Vocational Interest Blank-Strong Campbell Interest Inventory, UNIACT = Unisex Edition of the American College Testing, VPI = Vocational Preference Inventory, WPS = Work Preferences Survey, WSI = was-studiere-ich.de [what should I study]; ability measures: AFQT = Armed Forces Qualification Test, APM = Advanced Progressive Matrices, ASVAB = Armed Services Vocational Aptitude Scales, BAB = Ball Aptitude Battery, BIST = Berlin Intelligence Structure Test, DAT = Differential Aptitude Test, DAT-A = Differential Aptitude Tests-Adaptive, ETS KIT = Kit of Factor-Referenced Cognitive Tests, GATB = General Aptitude Test Battery, ISA = Intelligence-Structure-Analysis, KFT 4-12+R = Cognitive Ability Test-Revision, LPS-K = Leistungsprüfsystem-Short Version, PL-PQ = Australian Council of Educational Research Higher Test PL-PQ, WAIS-R = Wechsler Adult Intelligence Scale-Revised, WPT = Wonderlic Personnel Test. For detailed correlations of each study, see Appendix B.

Coding of primary studies

We coded the following data from each primary study: (a) full reference details, (b) study location, (c) year of publication, (d) year of data acquisition, (e) sample age, (f) gender distribution, (g) career level, (h) measured constructs, (i) reliability of constructs, and (j) correlations' effect size and direction.

Cognitive abilities. Cognitive ability (sub-)tests were classified using the CHC taxonomy. Thus, ability tests were coded to represent either general intelligence (*g*) or specific broad and narrow abilities. Detailed descriptions of each type of broad and narrow abilities as presented by Schneider and McGrew (2012) were used as a guideline for coding. Each ability test was coded independently by two of the authors. Few disagreements were discussed and resolved.

Gender. The gender distribution of the sample was identified. For moderator analyses, all-female, all-male, and those samples that provided correlation coefficients by gender were included, thus enabling comparisons between correlational patterns for females and males.

Age. The majority of studies reported a mean age of the sample (k = 27). Due to the small amount of studies that reported both mean age and correlations between interest types and specific cognitive abilities (k = 1 to k = 11), we decided to investigate age as a categorical moderator. For this purpose, samples with an average age below 20 years were compared to those with an average age of 20 years and older. This cut-off was chosen for theoretical reasons: Meta-analytical longitudinal research (Low et al., 2005) on change and stability of interests demonstrated that stability estimates from age 12 to age 18 (i.e., prior to graduation from high school) remained remarkably unchanged. However, during college years, interest stability increases dramatically. In the U.S., in Australia, and in Europe, individuals generally graduate from high school between 17 and 19 years of age. In order to compare high school and early adulthood samples, we split samples with an average age below 20 years from those with an average age of 20 years and older. Furthermore, samples showing great heterogeneity of age (SD > 5) were excluded from analyses (k = 4 with SD > 5 and k = 3 with missing SD_{age}).

Cohort. For each sample, we calculated an index for cohort by subtracting mean age from year of sample acquisition. If year of sample acquisition was not available, we used year of publication instead. Studies were subsequently assigned to one of six cohort groups: 1940s, 1950s, 1960s, 1970s, 1980s, and 1990s.

Statistical analyses

Analyses were conducted according to the validation generalization approach (Raju, Burke, Normand, & Langlois, 1991). This method is rooted in the meta-analytic approach by Hunter, Schmidt, and Jackson (1982) and corrects effect sizes individually for artifacts (i.e., sampling error, unreliability of measures) as opposed to using artifact distributions. Following recommendations by Hunter and Schmidt (2000), meta-analytic estimates were computed in a random-effects model using a software program by Raju and Fleer (2003). The fixed-effects model postulates that all included studies are homogeneous, sharing a common effect size, and all between-variance is caused by sample error, measurement error, or other adulterant or moderating influences. In contrast, the random-effects model allows the possibility for effect sizes to vary randomly from study to study.

Correlation coefficients were corrected for both sampling error and attenuation due to unreliability of both vocational interest and cognitive ability measures. Reliabilities were either obtained from the study or, if not reported, substituted by reliabilities stated in the test manuals. If neither approach was possible, we substituted scale reliabilities with a reliability estimate that was calculated based on the reliability information given in other studies for the specific construct.

Generally, only bivariate relationships between interests and cognitive ability measures that were drawn from three or more studies were retained for overall analyses. If studies reported two or more correlation coefficients for the same interest-ability-relation derived from one sample, these correlations were pooled. The reliability of the pooled predictors was estimated with Mosier's formula (Hunter & Schmidt, 2004). In accordance with Hunter and Schmidt (2004), we resigned from Fisher's *z*-transformations to pool correlations.

3.1.3 Results

We first present results for the interest-intelligence relation and then turn to the relation between interest themes and specific cognitive abilities. Gender, age, and cohort will be investigated as possible moderators. To interpret the magnitude of correlations (ρ), we applied Cohen's guideline: According to Cohen (1992), ρ = .10 is small, ρ = .30 is medium, and ρ = .50 is large. Correlations (ρ) are reported together with lower and upper bound of both 90% credibility value (CV) and 95% confidence interval (CI).

Holland's RIASEC types and general intelligence

We performed analyses for the RIASEC types and general intelligence. Results are shown in Table 2; if the 90% credibility value did not include zero, correlations are presented in bold. As expected, Investigative interests showed a positive correlation with *g* (ρ = .28, 95% oCI [.24, .33]), and this relation was the strongest for all interest types. However, we also found a small positive correlation with Realistic interests (ρ = .23, 95% CI [.17, .29]) and the 95% CIs for Investigative and Realistic interests overlapped. Further, results indicated a small negative correlation with Social interests (ρ = -.19, 95% CI [-.23, -.15]). For all above findings, neither the 90% CVs nor the 95% CIs included zero. The remaining correlations were close to zero. Thus, only Hypothesis 1a was supported. Noticeably, statistical artifacts (i.e., sampling error and measurement error) accounted for no more than 44% of variance in the correlations. Based on the 75%-rule by Hunter and Schmidt (2004), this indicates that the remaining variance is likely to be caused by additional artifacts that we have not yet taken into account, and moderator analyses are recommended.

Moderator analyses

Gender. Moderator analyses were conducted to determine whether the strength and direction of correlations between interests and general intelligence varied as a function of gender. In a first step, we included all studies that met the inclusion criterion. However, artifacts accounted for no more than 15% of variance in all correlation coefficients, thus indicating strong heterogeneity, while directions and magnitudes of ρ deviated considerably from those found for total samples. As highlighted by Kepes, McDaniel, Brannick, and Banks (2013), results and conclusions of meta-analyses can be heavily influenced by one or more effect sizes of deviant magnitude or by a single, large sample. In this case, Hunter and Schmidt

(2004) recommended a specific-sample-removed analysis, where meta-analytic results with and without excluded samples are compared to assess robustness of results. A close inspection of included studies in our meta-analysis revealed that data by Reeve and Heggestad (2004) deviated considerably both in magnitude and direction of correlation coefficients. Moreover, due to its large sample size, it strongly influenced the estimates of overall effect sizes. Thus, we decided to remove this sample from analyses and report results with and without data by Reeve and Heggestad (2004) to assess robustness of results. Moderator analyses including data from Reeve and Heggestad (2004) are reported in Appendix A.

As indicated in Table 3, when excluding data by Reeve and Heggestad (2004), we found positive correlations with Investigative (ρ_{males} = .22, 95% CI [.15, .29], $\rho_{females}$ = .23, 95% CI [.20, .27]) and Realistic interests (ρ_{males} = .11, 95% CI [.03, .19], $\rho_{females}$ = .20, 95% CI [.14, .25]) as well as a small negative correlation with Social interests (ρ_{males} = -.11, 95% CI [-.14, -.09], $\rho_{females}$ = -.15, 95% CI [-.17, -.13]) and *g* for both genders. However, for males the 90% CV for Realistic interests included zero. Relations between *g* and Artistic, Enterprising, and Conventional interests were negligible. With the exception of Conventional interests, all CIs for females and males overlapped. In sum, results by gender closely represented those found for total samples. Thus, the relationship between interests and general intelligence was not moderated by gender. However, mean variance accounted for by artifacts was larger for gender-specific analyses than for mixed samples (50.6% vs. 23.1%, respectively).

Cohort. Analyses for cohort as a continuous moderator were conducted using weighted multiple regression with inverse variance weights. Analyses were performed in SPSS using a module given by Wilson (2005). Results indicated that the relation between interests and *g* indeed varied by birth cohort. For males, the correlation between Realistic interests and *g* was positive for younger cohorts and negative for older cohorts (β = .72, p = .024), the reversed trend was found for Social interests (β = -.78, p = .035). For females, younger cohorts showed stronger positive relations between Realistic interests and *g* than older cohorts (β = .57, p = .027). For Enterprising interests, younger cohorts showed small negative relations with *g*, whereas the correlations were negligible for older cohorts (β = -.84, p = .021). In general, this moderator analysis must be interpreted with caution since the number of independent samples was distributed unequally among the cohorts. For the 1960s, only one study was available, whereas the 1970s and 1980s were overrepresented (*k* = 3 and *k* = 4, respectively).

Table 2. *Mean Effect Size Estimates and Confidence Intervals for the Correlations between Holland's RIASEC Types and General Intelligence*

	k	*N*	*r*	ρ	σ^2_ρ	% VE	90% CV	95% CI
Realistic	13	13,999	.20	**.23**	.010	9.9	[.10, .36]	[.17, .29]
Investigative	13	13,991	.25	**.28**	.005	17.8	[.20, .46]	[.24, .33]
Artistic	13	13,993	-.02	-.03	.006	16.5	[-.14, .08]	[-.07, .02]
Social	11	13,584	-.16	**-.19**	.004	21.0	[-.27, -.11]	[-.23, -.15]
Enterprising	12	13,909	-.07	**-.08**	.001	43.5	[-.13, -.03]	[-.11, -.05]
Conventional	12	13,908	.01	.01	.003	29.7	[-.06, .08]	[-.02, .05]

Note. *k* = number of independent samples; *N* = total sample size; *r* = sample size weighted mean correlation; ρ = estimated true score correlation (corrected for sample error and unreliability); $\sigma^2\rho$ = estimated variance for true score correlation; % VE = percentage of variance in ρ accounted for by statistical artefacts; 90% CV = lower and upper bound of the 90% credibility value for true score correlation; 95% CI = lower and upper bound of 95% confidence interval. Correlations are presented in boldface if the 90% credibility interval excludes zero.

Table 3. *Mean Effect Size Estimates and Confidence Intervals for the Correlations between Holland's RIASEC Types and General Intelligence by Sex (Excluding Data by Reeve & Heggestad, 2004)*

	k	*N*	*r*	ρ	σ^2_ρ	% VE	90% CV	95% CI
Males								
Realistic	8	6,072	.10	.11	.011	13.1	[-.03, .14]	[.03, .19]
Investigative	8	6,070	.19	**.22**	.003	34.8	[.15, .29]	[.17, .26]
Artistic	8	6,070	.03	.03	.002	43.5	[-.03, .06]	[-.01, .08]
Social	8	6,068	-.10	**-.11**	.000	100.0		[-.14, -.09]
Enterprising	8	6,068	-.08	**-.09**	.003	36.6	[-.11, -.02]	[-.14, -.04]
Conventional	8	6,069	-.05	**-.06**	.000	96.6	[-.07, -.05]	[-.09, -.03]
Females								
Realistic	8	7,183	.17	**.20**	.005	22.5	[.11, .29]	[.14, .25]
Investigative	9	7,326	.20	**.23**	.002	47.4	[.18, .28]	[.20, .27]
Artistic	8	7,179	.03	.04	.007	18.2	[-.07, .15]	[-.03, .10]
Social	8	7,178	-.13	**-.15**	.000	100.0		[-.17, -.13]
Enterprising	8	7,183	-.03	-.04	.002	49.3	[-.09, .01]	[-.08, .00]
Conventional	8	7,181	-.06	**.07**	.002	45.1	[.01, .13]	[.03, .11]

Note. *k* = number of independent samples; *N* = total sample size; *r* = sample size weighted mean correlation; ρ = estimated true score correlation (corrected for sample error and unreliability); $\sigma^2\rho$ = estimated variance for true score correlation; % VE = percentage of variance in ρ accounted for by statistical artefacts; 90% CV = lower and upper bound of the 90% credibility interval for true score correlation; 95% CI = lower and upper bound of 95% confidence interval. Correlations are presented in boldface if the 90% credibility interval excludes zero.

Holland's RIASEC types and specific cognitive abilities

In a next step, we performed analyses for the RIASEC types and narrow cognitive abilities. Analyses were conducted when data from a minimum of three independent samples were available (for descriptions of those narrow cognitive ability measures for which sufficient data were allocated, see Table 4). Some notations in the CHC framework differ from notations generally applied in cognitive ability research. To enhance interpretation of results and comparability of findings, we summarized findings on language development, quantitative reasoning, and visualization as findings on verbal, numerical, and spatial abilities, respectively. The results are presented in Table 5, and each of the six interest themes will be discussed. Again, correlations are marked in bold if the 90% credibility value did not include zero.

Small to moderate positive correlations were found between Realistic interests and spatial abilities (ρ = .34, 95% CI [.30, .40]), numerical abilities (ρ = .26, 95% CI [.18, .35]), and mechanical knowledge (ρ = .31, 95% CI [.23, .40]). All 90% credibility values excluded zero. Thus, Hypothesis 2a was supported. Further, a small positive correlation for Realistic interests and induction was revealed (ρ = .13, 95% CI [.08, .19], %VE = 100%). Due to an insufficient number of primary studies (*k* = 1) we were unable to investigate the relation between Realistic interests and motor coordination (Hypothesis 2b).

For Investigative interests, we found positive correlations with verbal (ρ = .21, 95% CI [.16, .27]), numerical (ρ = .25, 95% CI [.19, .30]), and spatial abilities (ρ = .27, 95% CI [.23, .31]). Furthermore, analyses revealed a small positive correlation with induction (ρ = .22, 95% CI [.14, .30]) and mechanical knowledge (ρ = .17, 95% CI [.02, .32]). All 90% credibility values excluded zero. Thus, Hypotheses 3a and 3b were supported.

As expected, we found a small positive correlation between Artistic interests and verbal abilities (ρ = .22, 95% CI [.18, .25]). Thus, Hypothesis 4 was supported. Further, analyses revealed a negative correlation with numerical abilities (ρ = -.18, 95% CI [-.24, -.12]).

For Social interests, we found negative relations with mechanical knowledge (ρ = -.28, 95% CI [-.37, -.19]) as well as with spatial (ρ = -.22, 95% CI [-.26, -.18]) and numerical abilities (ρ = -.21, 95% CI [-.27, -.14]). Thus, Hypothesis 5b was supported. Contrary to our expectation, the relation between Social interests and verbal abilities was very small negative (ρ = -.06, 95% CI [-.09, -.03]). Thus, Hypothesis 5a was not supported.

Correlations between Enterprising interests and narrow ability measures were negligible to small negative. Small negative correlations were found with spatial abilities (ρ = -.13, 95% CI [-.16, -.11]) and mechanical knowledge (ρ = -.14, 95% CI [-.16, -.12]). Deviant from expectation, we found no positive relation between Enterprising interests and verbal abilities (ρ = -.08, 95% CI [-.14, -.03]). Thus, Hypothesis 6 was not supported.

For Conventional interests, correlations with all narrower ability measures were negligible. Although the correlation with numerical abilities was positive (ρ = .08, 95% CI [.03, .12]), the 90% credibility interval included zero. Thus, Hypothesis 7a was not supported. In line with our hypotheses we found a positive albeit very small relation with perceptual speed (ρ = .06, 95% CI [.02, .13]). Thus, Hypothesis 7b was supported.

In sum, six out of eight hypotheses resulting from J. L. Holland's (1959, 1973, 1985) assumptions on the relationship between interest types and specific cognitive abilities were supported by our findings. Since statistical artifacts accounted for more than 75% of variance in only 9 out of 36 correlations between RIASEC themes and narrow ability measures, we conducted moderator analyses in a next step.

Table 4. *Classification of the Ability Measures: Broad and Narrow Cognitive Abilities and Definitions According to Schneider and McGrew (2012)*

Broad cognitive ability	Narrow cognitive ability	Definition
Verbal knowledge	Language development	General understanding of spoken language at the level of words, idioms, and sentences. *Core ability of verbal knowledge and crystallized intelligence* (g_c)
Fluid reasoning	Induction	The ability to observe a phenomenon and discover the underlying principles or rules that determine its behavior. *Core ability of fluid intelligence* (g_f)
	Quantitative reasoning	The ability to reason, either with induction or deduction, with numbers, mathematical relations, and operators.
Visual processing	Visualization	The ability to perceive complex patterns and mentally simulate how they might look when transformed (e.g., rotated, changed in size, partially obscured). *Core ability of visual processing (G_v)*
Processing speed	Perceptual speed	The speed at which visual stimuli can be compared for similarity or differences. *Core ability of processing speed (G_s)*
Domain specific knowledge	Mechanical knowledge	Knowledge about the function, terminology, and operations of ordinary tools, machines, and equipment.

Table 5. *Mean Effect Size Estimates and Confidence Intervals for the Correlations between Holland's RIASEC Types and Specific Cognitive Abilities*

	k	*N*	*r*	ρ	$σ^2_ρ$	% VE	90% CV	95% CI
Language Development (verbal ability)								
Realistic	14	10,097	.04	.05	.004	35.4	[-.03, .13]	[.01, .09]
Investigative	14	10,090	.17	**.21**	.008	18.6	[.10, .32]	[.16, .27]
Artistic	14	10,092	.17	**.22**	.003	40.1	[.15, .29]	[.18, .25]
Social	14	9,769	-.05	**-.06**	.001	76.7	[-.09. -.03]	[-.09, -.03]
Enterprising	14	10,094	-.07	-.08	.009	18.8	[-.20, .04]	[-.14, -.03]
Conventional	14	10,093	-.04	-.05	.005	31.1	[-.14, .04]	[-.09, -.01]
Induction								
Realistic	5	1,616	.10	**.13**	.000	100.0		[.08, .19]
Investigative	5	1,616	.16	**.22**	.004	58.6	[.14, .30]	[.14, .30]
Artistic	5	1,616	.05	.07	.009	38.5	[-.05, .19]	[-.03, .18]
Social	4	1,296	-.04	-.05	.001	80.6	[-.10, .00]	[-.13, .03]
Enterprising	5	1,616	-.09	-.12	.010	36.0	[-.25, .01]	[-.23, -.01]
Conventional	5	1,616	.04	.06	.009	39.6	[-.06, .18]	[-.05, .16]
Quantitative Reasoning (numerical ability)								
Realistic	10	9,076	.21	**.26**	.017	8.2	[.10, .42]	[.18, .35]
Investigative	10	9,068	.20	**.25**	.006	19.7	[.15, .35]	[.19, .30]
Artistic	10	9,070	-.14	**-.18**	.009	15.5	[-.30, -.06]	[-.24, -.12]
Social	10	9,067	-.17	**-.21**	.008	15.9	[-.33, -.09]	[-.27, -.14]
Enterprising	10	9,072	-.07	**-.08**	.001	65.1	[-.12, -.05]	[-.12, -.05]
Conventional	10	9,071	.06	.08	.004	27.9	[-.01, .17]	[.03, .12]
Visualization (spatial ability)								
Realistic	12	9,985	.28	**.34**	.006	21.2	[.25, .43]	[.30, .40]
Investigative	12	9,978	.21	**.27**	.003	35.2	[.19, .35]	[.23, .31]
Artistic	12	9,980	-.06	-.08	.006	23.8	[-.18, .02]	[-.13, -.03]
Social	12	9,977	-.17	**-.22**	.002	44.4	[-.28, -.16]	[-.26, -.18]
Enterprising	12	9,982	-.11	**-.13**	.000	81.5	[-.15, -.11]	[-.16, -.11]
Conventional	12	9,981	-.02	-.02	.001	61.8	[-.06, .02]	[-.05, .01]
Perceptual Speed								
Realistic	6	1,613	-.04	-.05	.014	29.4	[-.20, .10]	[-.16, .07]
Investigative	6	1,613	.06	.08	.002	76.6	[.03, .13]	[.01, .15]
Artistic	6	1,613	.05	.07	.012	32.8	[-.08, .22]	[-.04, .17]
Social	6	1,613	.03	.03	.003	63.4	[-.04, .10]	[-.05, .11]
Enterprising	6	1,613	**-.04**	-.06	.000	100.0		[-.09, -.02]
Conventional	6	1,613	**.06**	.08	.000	100.0		[.02, .13]
Mechanical Knowledge								
Realistic	3	992	.27	**.31**	.003	54.7	[.25, .37]	[.23, .40]
Investigative	3	992	.15	**.17**	.014	19.2	[.01, .33]	[.02, .32]
Artistic	3	992	-.11	-.12	.022	14.5	[-.31, .07]	[-.30, .06]
Social	3	992	-.24	**-.28**	.003	54.7	[-.35, -.21]	[-.37, -.19]
Enterprising	3	992	-.12	**-.14**	.000	100.0		[-.16, -.12]
Conventional	3	992	-.06	-.06	.000	93.4	[-.14, .01]	[-.14, .01]

Note. *k* = number of independent samples; *N* = total sample size; *r* = sample size weighted mean correlation; ρ = estimated true score correlation (corrected for sample error and unreliability); $\sigma^2\rho$ = estimated variance for true score correlation; % VE = percentage of variance in ρ accounted for by statistical artefacts; 90% CV = lower and upper bound of the 90% credibility interval for true score correlation; 95% CI = lower and upper bound of 95% confidence interval. Correlations are presented in boldface if the 90% credibility interval excludes zero.

Moderator analyses

Gender. As in previous analyses, correlation coefficients by gender were included in this moderator analysis to examine gender as a possible moderator of the relationship between interest types and narrow cognitive abilities. Correlations (ρ) are reported when at least three independent samples were available for this moderator analysis. Results are presented in Table 6; again, correlations in bold indicate that the 90% credibility interval excluded zero.

For *verbal abilities*, in line with previous findings, analyses showed a small positive correlation with Investigative (ρ = .19, 95% CI_{males} [.13, .26], 95% $CI_{females}$ [.12, .25], respectively) and Artistic interests (ρ = .23, 95% CI_{males} [.17, .30], 95% $CI_{females}$ [.18, .28], respectively) for females and males. Both 90% CVs and 95% CIs overlapped for both genders. All other correlations between interest types and verbal abilities were negligible for both genders.

For *numerical abilities*, we found a positive relation with Investigative interests (ρ_{males} = .15, 95% CI [.10, .21], $\rho_{females}$ = .17, 95% CI [.11, .23]) for both genders. Again, both 90% CVs and 95% CIs overlapped. For Realistic interests, albeit positive for both, the correlation with numerical abilities was stronger for females ($\rho_{females}$ = .19, 95% CI [.12, .27]) than for males (ρ_{males} = .07, 95% CI [-.03, .21]). However, the 95% CIs overlapped. Similarly, we found a small positive correlation with Conventional interests for females ($\rho_{females}$ = .10, 95% CI [.05, .15]) but not for males (ρ_{males} = .02, 95% CI [-.02, .06]). Again, however, 95% CIs overlapped slightly.

For *spatial abilities*, analyses showed a small positive correlation with Realistic (ρ_{males} = .25, 95% CI [.18, .33], $\rho_{females}$ = .27, 95% CI [.21, .33]) and Investigative interests (ρ_{males} = .21, 95% CI [.16, .26], $\rho_{females}$ = .22, 95% CI [.18, .26]), and small negative correlations with Social (ρ_{males} = -.16, 95% CI [-.22, -.10], $\rho_{females}$ = -.18, 95% CI [-.21, -.15]) and Enterprising interests (ρ_{males} = -.15, 95% CI [-.20, -.10], $\rho_{females}$ = -.11, 95% CI [-.14, -.09]) for males and females. All 90% CVs and 95% CIs overlapped. Correlations with Artistic and Conventional interests were overall negligible.

For *induction*, analyses indicated moderate positive relations with Investigative (ρ_{males} = .31, 95% CI [.15, .48], $\rho_{females}$ = .33, 95% CI [.24, .42]), and small negative relation with Enterprising interests (ρ_{males} = -.22, 95% CI [-.46, .03], $\rho_{females}$ = -.13, 95% CI [-.26, -.01]).

However, the 90% CV included zero for Enterprising interests. Again, the 95% CIs overlapped for males and females.

Thus, in general, the relationships between interests and specific cognitive abilities were not moderated by gender. However, mean variance accounted for by artifacts was larger for gender-specific analyses than for mixed samples: for verbal abilities 46.5% vs. 36.8%, for numerical abilities 53.9% vs. 25.4%, for spatial abilities 61.1% vs. 44.7%, and for induction 80.5% vs. 58.9%, respectively.

Age. Meta-analyses were conducted to determine whether the strength of correlations between interests and general intelligence varied as a function of sample age. Results of this moderator analysis are presented in Table 7. Unfortunately, we had to rely on a very small number of independent samples, especially for the younger age group with a mean age younger than 20 years (k = 2 to 3). With few exceptions, correlations between specific cognitive abilities and interests were slightly higher for the older age group.

For *verbal abilities*, we found a stronger positive relation with Investigative interests for the older age group (ρ = .31, 95% CI [.21, .41]) than for the younger group (ρ = .16, 95% CI [.15, .17]). Neither the 90% CVs nor the 95% CIs overlapped for either age group. Further, the relation between Enterprising interests was negative for the older age group (ρ = -.17, 95% CI [-.24, -.09]) but negligible for younger samples (ρ = -.01, 95% CI [-.03, .01]).

For *numerical abilities*, older samples showed stronger relations between interest types and numerical abilities for Realistic, Investigative, Artistic, and Social interests. Neither the 95% CIs nor the 90% CVs overlapped for either age group.

For *spatial abilities* and Realistic interests, we established stronger relations for older samples (ρ = .47, 95% CI [.44, .51]) than for younger samples (ρ = .29, 95% CI [.27, .31]). Neither the 90% CVs nor the 95% CIs overlapped for either age group.

Thus, results also indicated stronger relations for Investigative, Artistic, and Social interests for older samples. However, the CIs for older and younger samples slightly overlapped. With ρ = .47 for Realistic interests and numerical as well as spatial abilities, we found one of the highest correlations between any interest type and cognitive ability measures. Further, for 17 out of 36 correlations, statistical artifacts explained 100% of variance in the correlation, thereby indicating homogeneity of the relations. Thus, we found evidence for age-specific differences in the relation between vocational interests and specific cognitive abilities.

Cohort. As for general intelligence, results indicated that the relation between interests and specific cognitive abilities partly varied by birth cohort. For verbal abilities, older male cohorts showed small negative correlations with Social interests, whereas the correlations were negligible for younger male cohorts (β = .68, p = .020). For numerical abilities, the relation with Investigative interests for females was negative for more recent cohorts but positive for older cohorts (β = -.70, p = .006); the reversed trend was found for Conventional interests (β = .76, p = .013). For spatial abilities, analyses indicated that for males, younger cohorts showed a stronger negative relation with Enterprising interests than older cohorts (β = -.68, p = .016). For Conventional interests, results indicated that whereas for males the relation with spatial abilities was negative in younger cohorts, but positive in older cohorts (β = -.67, p = .006), the reversed trend was found for females (β = .75, p = .021). Moreover, for females, the relation with Realistic interests became stronger in more recent cohorts (β = .53, p = .031). Again, this moderator analysis must be interpreted with caution since the number of independent samples was small and distributed unequally among the cohorts. For the 1960s, 1970s, and 1990s, a maximum of two studies was available, whereas the 1980s were slightly overrepresented ($2 \leq k \leq 4$).

Follow-up analysis

Publication bias is a possible danger to the validity of any meta-analysis. Thus, we investigated the presence of potential bias against small or non-significant findings with a funnel graph that plots sample sizes against effect sizes. From visual inspections of the plot, no exclusion of small or negative results from small samples was detectable. It should be mentioned that many correlation matrices were drawn from studies that did not explicitly investigate the relation between vocational interests and cognitive abilities. Thus, withdrawing from reporting non-significant correlations between both measures had not been a problem in these studies.

Table 6. *Mean Effect Size Estimates and Confidence Intervals for the Correlations between Holland's RIASEC Types for Selected Specific Abilities by Sex*

	k	*N*	*r*	ρ	σ^2_ρ	% VE	90% CV	95% CI	*k*	*N*	*r*	ρ	σ^2_ρ	% VE	90% CV	95% CI
	Males								Females							
Language Development (verbal ability)																
Realistic	10	4,344	-.04	-.05	.008	30.3	[-.17, .07]	[-.11, .02]	10	5,007	.04	**.05**	.001	74.9	[.01, .09]	[.01, .09]
Investigative	10	4,342	.16	**.19**	.008	28.0	[.08, .30]	[.13, .26]	10	5,002	.15	**.19**	.008	26.3	[.07, .31]	[.12, .25]
Artistic	10	4,342	.18	**.23**	.006	35.0	[.13, .33]	[.17, .30]	10	5,004	.18	**.23**	.003	46.0	[.16, .30]	[.18, .28]
Social	10	4,340	-.02	-.03	.002	64.4	[-.09, .03]	[-.07, .02]	10	5,003	-.05	**-.06**	.000	100.0		[-.09, -.03]
Enterprising	10	4,340	-.03	-.04	.010	23.7	[-.17, .09]	[-.11, .03]	10	5,008	-.04	-.05	.008	26.8	[-.17, .07]	[-.12, .02]
Conventional	10	4,341	-.05	-.07	.003	57.7	[-.14, .00]	[-.12, -.02]	10	5,006	-.02	-.02	.004	44.7	[-.10, .06]	[-.07, .03]
Quantitative Reasoning (numerical ability)																
Realistic	7	4,033	.06	.07	.018	12.3	[-.10, .24]	[-.03, .18]	8	4,809	.16	**.19**	.009	20.2	[.07, .31]	[.12, .27]
Investigative	7	4,031	.12	**.15**	.003	46.1	[.08, .22]	[.10, .21]	8	4,803	.14	**.17**	.005	29.8	[.08, .26]	[.11, .23]
Artistic	7	4,031	-.07	**-.09**	.001	74.9	[-.13, -.05]	[-.13, -.04]	8	4,805	-.09	**-.11**	.005	35.4	[-.19, -.03]	[-.17, -.06]
Social	7	4,029	-.10	**-.12**	.000	100.0		[-.14, -.10]	8	4,804	-.10	**-.13**	.003	44.1	[-.20, -.06]	[-.18, -.08]
Enterprising	7	4,029	-.05	-.06	.005	35.0	[-.15, .03]	[-.13, .00]	8	4,809	-.04	**-.05**	.000	100.0		[-.08, -.03]
Conventional	7	4,030	.02	**.02**	.000	95.3	[.01, .03]	[-.02, .06]	8	4,807	.08	**.10**	.002	53.7	[.04, .14]	[.05, .15]
Visualization (spatial ability)																
Realistic	9	3,673	.20	**.25**	.008	29.0	[.14, .36]	[.18, .33]	10	4,925	.21	**.27**	.007	30.1	[.16, .38]	[.21, .33]
Investigative	9	3,671	.17	**.21**	.003	58.2	[.15, .27]	[.16, .26]	10	4,920	.17	**.22**	.001	69.0	[.17, .27]	[.18, .26]
Artistic	9	3,671	-.03	-.04	.003	57.4	[-.11, .03]	[-.09, .02]	10	4,922	.02	.03	.007	32.8	[-.08, .14]	[-.03, .09]
Social	9	3,669	-.12	**-.16**	.005	44.9	[-.25, -.07]	[-.22, -.10]	10	4,921	-.14	**-.18**	.000	100.0		[-.21, -.15]
Enterprising	9	3,669	-.12	**-.15**	.002	62.2	[-.21, -.09]	[-.20, -.10]	10	4,926	-.09	**-.11**	.000	100.0		[-.14, -.09]
Conventional	9	3,670	-.06	-.08	.003	53.5	[-.16, .00]	[-.13, -.02]	10	4,924	.01	.02	.000	95.8	[.00, .04]	[-.02, .06]
Induction																
Realistic	3	284	.04	.06	.005	79.4	[-.03, .15]	[-.12, .25]	3	333	.02	**.03**	.000	100.0		[-.12, .18]
Investigative	3	284	.22	**.31**	.005	78.6	[.23, .39]	[.15, .48]	3	333	.23	**.33**	.000	100.0		[.24, .42]
Artistic	3	284	.12	.18	.030	40.5	[-.04, .40]	[-.07, .43]	3	333	.07	.09	.013	56.8	[-.05, .16]	[-.10, .29]
Social	3	284	-.07	-.10	.024	44.8	[-.30, .10]	[-.33, .14]	3	333	-.05	**-.07**	.000	100.0		[-.21, .06]

Enterprising	3	284	-.15	-.22	.029	39.1	[-.44, .01]	[-.46, .03]	3	333	-.09	**-.13**	.000	100.0		[-.26, -.01]
Conventional	3	284	.01	.01	.029	41.5	[-.21, .23]	[-.24, .26]	3	333	-.02	-.03	.048	26.3	[-.31, .25]	[-.32, .26]

Note. k = number of independent samples; N = total sample size; r = sample size weighted mean correlation; ρ = estimated true score correlation (corrected for sample error and unreliability); $\sigma^2\rho$ = estimated variance for true score correlation; % VE = percentage of variance in ρ accounted for by statistical artefacts; 90% CV = lower and upper bound of the 90% credibility interval for true score correlation; 95% CI = lower and upper bound of 95% confidence interval. Correlations are presented in boldface if the 90% credibility interval excludes zero.

Table 7. *Mean Effect Size Estimates and Confidence Intervals for the Correlations between Holland's RIASEC Types and Specific Cognitive Abilities by Age*

	k	*N*	*r*	ρ	σ^2_ρ	% VE	90% CV	95% CI	*k*	*N*	*r*	ρ	σ^2_ρ	% VE	90% CV	95% CI
	Age (< 20 years)								Age (≥ 20 years)							
Language Development (verbal ability)																
Realistic	3	5,634	.02	.02	.002	34.7	[-.03, .07]	[-.03, .08]	4	2,440	.11	**.13**	.000	92.7	[.11, .15]	[.08, .18]
Investigative	3	5,627	.13	**.16**	.000	94.6	[.15, .17]	[.13, .19]	4	2,440	.25	**.31**	.008	21.5	[.20, .42]	[.21, .41]
Artistic	3	5,629	.17	**.22**	.000	100.0		[.20, .24]	4	2,440	.12	**.15**	.005	33.6	[.07, .23]	[.07, .24]
Social	3	5,306	-.06	**-.07**	.000	100.0		[-.10, -.04]	4	2,440	-.07	**-.10**	.001	79.3	[-.14, -.06]	[-.15, -.04]
Enterprising	3	5,631	-.01	-.01	.000	77.8	[-.03, .01]	[-.04, .03]	4	2,440	-.13	**-.17**	.004	39.0	[-.25, -.09]	[-.24, -.09]
Conventional	3	5,630	-.02	-.03	.002	28.7	[-.09, .03]	[-.09, .03]	4	2,440	-.03	-.03	.005	32.3	[-.12, .06]	[-.12, .05]
Quantitative Reasoning (numerical ability)																
Realistic	2	5,310	.18	**.22**	.000	100.0		[.20, .25]	3	2,221	.38	**.47**	.004	29.9	[.39, .55]	[.39, .55]
Investigative	2	5,302	.18	**.23**	.000	100.0		[.21, .24]	3	2,221	.28	**.35**	.000	98.5	[.34, .36]	[.30, .39]
Artistic	2	5,304	-.13	**-.16**	.003	18.5	[-.22, -.10]	[-.24, -.08]	3	2,221	-.23	**-.30**	.002	57.7	[-.35, -.25]	[-.37, -.24]
Social	2	5,301	-.16	**-.20**	.000	100.0		[-.21, -.18]	3	2,221	-.26	**-.34**	.000	100.0		[-.38, -.30]
Enterprising	2	5,306	-.08	**-.10**	.000	100.0		[-.13, -.07]	3	2,221	-.03	**-.04**	.000	100.0		[-.09, .01]
Conventional	2	5,305	.03	.04	.001	40.9	[.00, .08]	[-.01, .09]	3	2,221	.09	**.12**	.005	30.0	[.03, .21]	[.02, .21]
Visualization (spatial ability)																
Realistic	3	6,000	.23	**.29**	.000	100.0		[.27, .31]	4	2,440	.38	**.47**	.000	100.0		[.44, .51]
Investigative	3	5,993	.20	**.25**	.001	41.2	[.21, .29]	[.20, .30]	4	2,440	.25	**.31**	.000	100.0		[.29, .34]
Artistic	3	5,995	-.04	-.06	.003	23.0	[-.13, .01]	[-.12, .01]	4	2,440	-.14	**-.18**	.002	53.2	[-.24, -.12]	[-.25, -.11]
Social	3	5,992	-.17	**-.21**	.000	100.0		[-.24, -.18]	4	2,440	-.23	**-.29**	.000	100.0		[-.33, -.24]
Enterprising	3	5,997	-.10	**-.13**	.000	100.0		[-.15, -.10]	4	2,440	-.11	**-.14**	.000	100.0		[-.16, -.11]
Conventional	3	5,996	-.04	**-.05**	.000	100.0		[-.07, -.03]	4	2,440	.03	**.03**	.000	100.0		[.01, .05]

Note. *k* = number of independent samples; *N* = total sample size; *r* = sample size weighted mean correlation; ρ = estimated true score correlation (corrected for sample error and unreliability); $\sigma^2\rho$ = estimated variance for true score correlation; % VE = percentage of variance in ρ accounted for by statistical artefacts; 90% CV = lower and upper bound of the 90% credibility interval for true score correlation; 95% CI = lower and upper bound of 95% confidence interval. Correlations are presented in boldface if the 90% credibility interval excludes zero.

3.1.4 Discussion

Summary

The main goal of the present meta-analysis was to investigate the relation between Holland’s RIASEC themes and cognitive abilities. Specifically, we were interested in whether (a) the relation between interests and cognitive abilities varies by gender, (b) the relation between interests and cognitive abilities becomes more pronounced by age, and (c) the findings from Ackerman and Heggestad’s (1997) review could be supported by our quantitative analyses. We analyzed results from 27 primary studies (and 29 independent samples) and believe that our findings provide important insights into the relation between vocational interests and cognitive abilities.

Our results support the notion of small to medium correlations between vocational interests and cognitive abilities. In accordance with J. L. Holland’s (1959, 1973, 1985) assumptions, we found (a) positive relations between Realistic interests and mechanical knowledge, (b) positive relations between Investigative interests and *g* as well as numerical abilities and induction, (c) positive relations between Artistic interests and verbal abilities, (d) negative relations between Social interests and mechanical knowledge, and (e) positive albeit very small relations between Conventional interests and perceptual speed. Deviant from Holland’s assumptions, we found negative relations between Enterprising and Social interests and verbal abilities. We further established (a) positive relations between Realistic interests and spatial as well as numerical abilities, and (b) positive relations between Investigative interests and spatial as well as verbal abilities. All findings are in accordance with Ackerman and Heggestad’s (1997) review. In sum, whereas Realistic, Investigative, and Artistic interests were linked to cognitive abilities, we found only negligible or negative relations for Enterprising and Social interests and specific cognitive abilities. Armstrong et al. (2011) pointed out two reasons to account for these findings: First, this may reflect a lack of traditional cognitive ability measures to map abilities used to work effectively with others (i.e., social or management skills) or second, cognitive abilities may not be critical for job performance in environments that strongly emphasize interpersonal relations. However, in her work on an occupational aptitude patterns map, L. S. Gottfredson (1986) assigned at least average levels of *g*, verbal, and numerical abilities to jobs that involved dealing with social and economic relations. In sum, interests and cognitive abilities were found to be modestly correlated with cognitive abilities. This limited overlap between interests and cognitive abilities suggests that the assessment of each individual difference measure provides unique information about an individual. Thus, neither measure should be replaced by the other in both research and practice[1].

Past research suggested that gender moderates the relation between RIASEC types and cognitive abilities (Carless, 1999; Reeve & Heggestad, 2004). However, examining gender as a moderator, we found that the direction and magnitude of correlations between vocational interests and cognitive abilities were comparable for females and males. Thus, counter to expectations, the relation between cognitive abilities and interests was not moderated by gender. Likewise, gender was not found to be a substantial moderator of the relation between interests and personality (Staggs et al., 2007).

Analyzing birth cohorts from the 1940s to 1990s, moderator analyses indicated some cohort effects for the relation of interests and cognitive abilities. Cohort effects in vocational interests are in general attributed to changes in the labor market. First, there has been a steady increase in the number of women entering the workforce as well as an increase in college and graduate degrees earned by women (Bubany & Hansen, 2011). Second, there has been a general decline in individuals working in the Realistic area (occupations such as auto mechanic, aircraft controller, surveyor, or farmer) and a steady increase in individuals employed in the Enterprising area (i.e., occupations such as business executive, salesperson, supervisor, and manager) as reported by Reardon, Bullock, and Meyer (2007). Third, there has been a pronounced decline in manual and cognitive routine tasks and a marked increase in complex cognitive tasks, such as planning, selling, and doing research, in recent decades particularly due to technological changes (Spitz-Oener, 2006). Altogether, these shifts in the labor market may help explain why we found indications of cohort effects in our analyses. Our results, however, must be interpreted with caution since we relied on very few samples

[1] In vocational counseling instead of focusing on the level of a particular type of interest, individuals are characterized by a two- or three-letter code (i.e., by the two or three interest types that resemble the person most in descending order). Educational and occupational environments are similarly characterized. By matching an individual's three-letter code to occupational characteristics potential career choices are identified.

for these moderator analyses. Further, samples were distributed unequally among the decades with considerably more samples from the 1970s and 1980s.

Cognitive investment theories (Cattell, 1987; Schmidt, 2011) propose that personality and interests guide the development of crystallized intelligence, specifically the acquisition of knowledge, skills, and aptitudes. Recently, Von Stumm and Ackerman (2013) found that general interest in knowledge acquisition is positively correlated with crystallized intelligence, academic performance, and acquired knowledge. We examined age as potential moderator and hypothesized that we would find stronger relations between specific cognitive abilities that are highly influenced by experience and knowledge acquisition, that is, measures of crystallized intelligence and related vocational interests. Within the CHC framework, domain-specific knowledge, language development, and quantitative knowledge are perceived as acquired knowledge constructs (Schneider & McGrew, 2012). For both language development as well as quantitative knowledge, we indeed found evidence for more pronounced relations with various interest types for older samples than for younger samples.

In summary, our findings lead to three main conclusions: (a) we found empirical support for small to moderate correlations between vocational interests and cognitive abilities providing evidence for Holland's assumptions on the relation between interest types and cognitive abilities, (b) deviant from past research, we established that relations between interests and cognitive abilities were comparable for females and males, and (c) we further found support for the notion that the relation between vocational interests and specific cognitive abilities, especially those that are influenced by experience and knowledge acquisition, becomes more pronounced with age.

Limitations

First, due to the comparably small number of studies that reported correlation coefficients between vocational interests and specific cognitive abilities as well as information on mean sample age, we were unable to consider age as a continuous moderator and instead relied on group comparisons.

Second, by focusing on Holland's RIASEC framework, we indeed based our analyses on the most prevalent model of vocational interests. Nevertheless, we had to exclude studies that relied on other theoretical frameworks such as basic interest markers. Basic interest scales measure interests on a lower level of generality than Holland's RIASEC framework such as interests in specific fields of work (e.g., engineering, teaching, physics, and theology).

Third, we decided upon the CHC framework as a classification system for our analyses since this taxonomy of cognitive abilities is widely accepted and empirically well validated (Alfonso et al., 2005). Further, it enabled us to classify the diverse specific cognitive ability measures administered in the primary studies. However, our results may be influenced by our choice of cognitive ability taxonomy since the CHC theory was not developed to implement relations among cognitive abilities and other individual difference measures such as vocational interests or personality.

Fourth, although primary studies offered a wide range of cognitive ability measures, we focused our analyses on a small selection of specific cognitive abilities due to the small number of primary studies. The majority of primary studies have been conducted in the field of career counseling. Measures of specific abilities are especially important in career counseling since they help individuals match their educational and occupational choices to their individual constellation of abilities (Humphreys et al., 1993). In this setting, measures of verbal, numerical, and spatial abilities are often administered and are therefore overrepresented in our meta-analysis. However, especially the Artistic type is associated with abilities such as divergent thinking and artistic abilities that are not captured by our meta-analysis. The same accounts for social or interpersonal skills and management abilities which are associated with the Social and the Enterprising type.

Lastly, the current study did not correct for range restriction which is likely to affect cognitive ability measures rather than vocational interest measures. Several primary studies included in this meta-analysis were based on college or university samples where admission is usually based on aptitude test scores. Thus, it is possible that our meta-analytic correlation coefficients are underestimated because of range restriction.

Implications for future research

As Lowman and Carson (2013) highlighted, it may be important to investigate not only the relationship between specific cognitive abilities and vocational interests but also between *g*-free specific abilities and vocational interests. Generally, measures of specific cognitive abilities (i.e., measures of broad and narrow cognitive abilities) correlate strongly with measures of general intelligence (*g*). Thus, to get an accurate understanding of the relation between specific cognitive abilities and interests, correlations should be controlled for *g*. Both Carson (1996, 1998a) as well as Pässler (2011) showed that the relation between Holland's RIASEC types and specific cognitive abilities alters once *g* is controlled for. Overall, the correlation between vocational interests and specific cognitive abilities is considerably reduced. Pässler (2011) reported negligible to small relations between Investigative interests and verbal, numerical, and spatial abilities when controlling for *g*. Further, the negative correlation between Social interests and numerical and spatial abilities diminished considerably. Thus, variability of past research on the relation between interests and cognitive abilities may be partly attributed to not considering differences in samples' *g*-level. Most research on the relationship between vocational interest and cognitive abilities relies on either high school students applying for college or college students. However, college samples generally display above average *g*-levels. This notion is important since research showed that at higher levels of *g*, specific cognitive abilities become more differentiated–generally referred to as Spearman's *law of diminishing returns* or *differentiation hypothesis* (for a comprehensive review, see Deary et al., 1996). Further, individuals with higher levels of intelligence were found to show broader vocational interests (Johnson & Bouchard, 2009). Broad interests and high levels of general intelligence may lead to crystallization of intelligence and knowledge acquisition in a wide range of content areas. Thus, samples' *g*-level may influence the correlational pattern of interests and specific abilities.

As summarized, we established negligible relations only between both Enterprising and Social interests and cognitive ability measures. As proposed by Armstrong et al. (2011), this may reflect a lack of traditional cognitive ability measures to map abilities used to work effectively with others such as social or emotional intelligence. Investigating the relation between Holland's RIASEC types and social intelligence, Lowman and Leeman (1988), for example, found that social and interpersonal skills were positively related with Enterprising but not Social interests. However, as pointed out by Mackintosh (2011), measures of social and emotional competence tend to show only moderate correlations with traditional measures of intelligence and are best characterized as a blend of both ability and personality. Thus, further research is needed to establish how measures of social and emotional competencies can be integrated in a framework of interests, cognitive abilities, and personality.

Finally, although we found indication that the relation between interests and specific cognitive abilities becomes more pronounced with age, our meta-analysis was based only on cross-sectional data. Thus, no causal inferences can be drawn. To further investigate the question whether interests guide the development of crystallized intelligence (i.e., the acquisition of knowledge, skills, and aptitudes), analyses of longitudinal data are necessary. For such studies, we recommend the investigation of possible moderators such as specialization in education and reinforcement as well as deprivation and discouragement during socialization.

3.2 Gender-related differential validity and differential prediction in interest inventories

3.2.1 Aims and hypotheses

Aims

Research on prediction bias in interest inventories has merely relied on comparing hit rates for females and males or evaluating gender differences in congruence indices. Nevertheless, both NIE and APA Standards clearly emphasize the suspension of gender differences by testing for differential prediction. Several meta-analyses have shown relationships between person-environment fit and occupational satisfaction (Assouline & Meir, 1987; Tranberg, Slane, & Ekeberg, 1993; Tsabari et al., 2005). Following the Standards, we first compared validity coefficients for the prediction of person-environment fit and satisfaction for females and males to test for predictive validity, and second, we examined gender differences in the slopes and intercepts in a regression model predicting person-environment fit to test for differential prediction.

Furthermore, by following Meade and Tonidandel's (2010) suggestions, we combined an internal and external approach for establishing test fairness. As discussed above, research (Aros et al., 1998; Einarsdóttir & Rounds, 2009) shows that by eliminating interest items showing large DIF, gender differences can be reduced. Both Aros et al. (1998) and Einarsdóttir and Rounds (2009) ascribed the observed gender-related DIF to a sex-type dimension influencing female and male responses. Nevertheless, neither Aros et al. (1998) nor Einarsdóttir and Rounds (2009) compared instrument validity of the SII with its DIF-optimized version. Evidence from UNIACT research suggests that accuracy of classifying persons into their occupational preference group was comparable for one version of the instrument where gender differences were eliminated on the item level and a traditional version of the same instrument (Prediger & Lamb, 1979). Thus, in a second step, we compared the validity of a traditional constructed interest inventory with its DIF-optimized version and analyzed both instruments for differential validity and differential prediction as well as compared findings.

The study is based on the following publication:
Pässler, K., Beinicke, A., & Hell, B. (2014). Gender-related differential validity and differential prediction in interest inventories. *Journal of Career Assessment*, *22*(1), 138-152. doi:10.1177/1069072713492934

Hypotheses

According to J. L. Holland (1997), individuals tend to gravitate toward environments that are consistent with their interest profile. Furthermore, the level of person-environment congruence will increase the likelihood of subjective person-environment fit as well as satisfaction with choice. Thus, to demonstrate instrument fairness neither C-Index level (Hypothesis 1) nor correlation between C-Index and subjective person-environment fit (Hypothesis 2) nor satisfaction (Hypothesis 3) should differ between females and males. Moreover, when predicting person-environment fit by an individual's interest score the slopes and intercepts of the regression model should not differ between females and males (Hypothesis 4).

Furthermore, we were interested in the question whether eliminating interest items showing large DIF influences instrument validity. Therefore, we additionally examined a DIF-optimized version of an interest inventory for differential validity and differential prediction and compared these findings to those of the standard version.

3.2.2 Methodology

Participants

The sample comprised 797 students (62.8% women and 37.2% men). Due to unrealistic response patterns (total processing time of the questionnaire less than 20 min), a reduced sample size of *N* = 736 was included in further analyses. Participants were either enrolled in vocational training (50.8%) or university programs (49.2%). In order to guarantee sufficient heterogeneity of job-related interests, we investigated different fields of vocational training and fields of study. That is, the assigned three-letter Holland codes vary. Thus, we investigated three groups of trainees during their apprenticeship (vocational training): digital media designers (15.1%), hotel managers (18.7%), and design draftspeople (17.0%). Students were recruited from university, studying different majors such as chemistry (15.5%), economics (17.5%), or linguistics (16.2%). Table 8 shows characteristics of the sample separated by fields of studies.

We tried to balance the number of female and male trainees and students in the sample. However, depending on the field of study, the distribution of gender in the population is inhomogeneous, favoring females for subjects with a stronger focus on social interaction and males for more technical subjects. Participant's age ranged from 16 to 48 years (*M* = 21.42, *SD* = 2.89). Data were collected online during a 6-month period.

Table 8. *Characteristics of the Sample Separated by Field of Studies*

Abbreviations	Field of Study	GIST-R Three-Letter Code	n_{female}	n_{male}	n_{total}
MD	Digital media design	AER	72	39	111
HM	Hotel management	ECS	112	26	138
ED	Engineering drawing	RIE	55	70	125
C	Chemistry	IRC	47	67	114
E	Economics	ECI	72	57	129
L	Linguistics	AIS	104	15	119

Note. GIST-R = Revised General Interest Structure Test. The GIST-R three-letter code is taken from the GIST-R Manual. *N* = 736.

Instruments/Measures

Revised General Interest Structure Test (GIST-R). Participant's occupational interests were assessed using the GIST-R (''Allgemeiner Interessen-Struktur-Test''; Bergmann & Eder, 2005). The GIST-R is a widely used German interest inventory based on J. L. Holland's (1959, 1997) RIASEC model. Participants were asked to rate their individual level of interest in the represented activity on a 5-point Likert-type scale, ranging from 1 (I am not interested in this at all; I do not enjoy doing this at all) to 5 (I am very interested in this; I enjoy doing this very much). Each participant's score on the 60 items (10 per dimension) was aggregated to the six RIASEC dimensions. Reliabilities (Cronbach's αs) for the GIST-R scales range between .82 and .87; the 1-month retest reliability ranges from r = .85 to r = .92 (Bergmann & Eder, 2005). The GIST-R is the best validated interest inventory in the German-speaking countries (i.e., Germany, Austria, and Switzerland). Scale score correlations between the GIST-R and matching scales from an adaptation of Holland's SDS instrument (Jörin, Stoll, Bergmann, & Eder, 2003) range from r = .60 to r = .75. Furthermore, evidence for the structural validity of the GIST-R is provided (Nagy et al., 2010).

Person-Environment Fit Score/Satisfaction. Students indicated their perceived person-environment fit with their chosen training course or field of study on a scale ranging from 1 (*not at all*) to 6 (*perfectly*) and furthermore rated their satisfaction with their chosen training course or field of study on a scale ranging from 1 (*very dissatisfied*) to 6 (*very satisfied*). Albeit having several psychometric shortcomings, assessing general job satisfaction by a single-item measure is well established in the literature. In their meta-analysis, Wanous, Reichers, and Hudy (1997) attest this approach satisfying reliability and construct validity. Correlation between perceived person-environment fit and satisfaction was .42.

Cognitive Ability. We assessed participants' cognitive ability to control for possible moderator effects. Cognitive ability was chosen as a moderator since individuals evaluation of their perceived person-environment fit as well as satisfaction with occupational choice might rely on both the perceived fit between an individual's interest and the occupational requirements as well as on the perceived fit between an individual's cognitive abilities and the occupational demands. For assessing cognitive ability, a short version of an ability test designed for vocational counseling purposes (for details, see Hell, Pässler, & Schuler, 2009) measuring verbal, numerical, and spatial abilities was administered focusing on the dimensions: verbal classifications (Cronbach's α = .73; 7 items), verbal analogies (.62; 6), numerical sequences (.67; 4), rule of three (.57; 3), quantitative comparisons (.64; 4), surface development (.54; 4), and mental rotations (.66; 3). Then, we aggregated these dimensions to an overall cognitive ability score for each participant. Instrument development was conducted according to the Berlin model of intelligence structure (Jäger, 1984; for details, see Pässler & Hell, 2012). Positive evidence for instrument validity for predicting content-specific high school and college grades as well as college major choice (Pässler & Hell, 2012).

Procedures

DIF-Optimized Version. Analyses of DIF were carried out to investigate whether GIST-R items assess the same underlying constructs for both women and men. Two approaches were pursued: nonparametric and parametric methods. First, a nonparametric method for testing polytomous items for DIF based on calculating the Liu-Agresti cumulative common log-odds ratio (L-A LOR; Liu & Agresti, 1996) was applied. This method is implemented in DIF analysis system 5.0 (Penfield, 2009) and is based on contingency tables. Derived from the Mantel-Haenszel common odds ratio used for dichotomous items, the L-A LOR is seen as its generalization for polytomous items (Penfield & Algina, 2006). In order to evaluate the DIF-size, we followed the classification system by Educational Testing Service (Zieky, 1993). Three categories were found: A for items with negligible DIF (L-A LOR < .43), B for items with slight to moderate DIF (.43 ≤ L-A LOR < .64), and C for items with moderate to large DIF (L-A LOR ≥ .64). Second, parametric DIF analyses were conducted with ConQuest (Wu, Adams, Wilson, & Haldane, 2007). Afterward, the ConQuest DIF estimates were transformed to allow comparisons to L-A LOR values. Both methods are precisely described by Wetzel and Hell (2013). Although the Educational Testing Service recommends eliminating both items showing slight to moderate (B) and large (C) DIF, we decided to eliminate only C items else an insufficient number of items would have been left in the Realistic domain. This procedure led to a reduced number of items in all but one RIASEC dimension in the DIF-optimized inventory (for details, see Wetzel & Hell, 2013): Realistic (5 items), Investigative (7), Artistic (8), Social (10), Enterprising (6), and Conventional (8).

Congruence. Students were asked to indicate their current training course or field of study. Each educational choice was assigned a three-letter RIASEC code, according to the register of occupational codes in the GIST-R manual by Bergmann and Eder (2005). Congruence was then calculated between a participant's three-letter code and the RIASEC codes of the current training course or field of study using the C-Index (Brown & Gore, 1994). The C-Index ranks from 0 to 18 with higher scores indicating higher congruence between a participant's vocational interests and educational choice. Separate C-Indices were calculated for the standard and the DIF-optimized GIST-R scales.

Differential Prediction. Differential prediction is generally assessed within a moderated multiple regression (MMR) framework. We used MMR as an inferential procedure to compare two different least squares regression equations (Aiken & West, 1991). MMR establishes whether a moderating variable, such as gender, influences the predictor-criterion relationship (Aguinis, 2004). The MMR model includes the first-order effects for predicting Y (quantitative dependent variable) from X (predictor), Z (a second binary predictor), and the product term X x Z which carries information regarding the moderating effect of Z.

$$Y = a + b_1X + b_2Z + b_3XZ + e.$$

If the interaction term is significant, the predictive relationship with the criterion (i.e., the slope) differs across subgroups defined by the biasing variable such as gender. Enhancing the interpretation and plotting of simple slopes, all continuous first-order effects including control variables were mean centered. Categorial variables were centered using dummy coding. Interactive terms were computed using the mean-centered first-order terms. Unstandardized regression coefficients are reported since standardized solutions are afflicted with difficulties when a product term is involved in the model (Jaccard & Turrisi, 2003).

Furthermore, when applying MMRs, concerns of their statistical power are often raised (Aguinis, Culpepper, & Pierce, 2010; Aguinis & Stone-Romero, 1997). Following the suggestions of Aguinis et al. (2010), we referred to a conservative significance level of $p < .10$ for the interpretation of the interaction effect. By adopting this conservative criterion, we decided that mistakenly concluding no bias when the data suggest bias is present (Type II error) would be more severe than concluding bias when the data suggest otherwise (Type I error).

For all MMRs, we entered the mean-centered interest variable (regarding the dominant letter specifically for each training course or field of study) and gender in Step 1 of the regression equation. The product term between interest and gender was additionally placed in Step 2. As Aguinis (2004) advises, the first-order effect was entered first and the product term second. Person-environment fit was used as dependent variable in all MMR models.

3.2.3 Results

DIF Optimization

The DIF-optimized version leads to a reduction of mean differences between women and men regarding the dimensions R, I, and A, but not for S and E (see Table 9). No significant gender differences were found for dimension C. For further details regarding the comparison between the DIF-optimized and standard tests, see Wetzel and Hell (2013).

Table 9. *Reduction of Gender-Specific Mean Differences in the DIF-Optimized Version Regarding the RIASEC Dimensions*

	GIST-R					DIF-GIST-R					
	Female (*N* = 462)		Male (*N* = 274)			Female (*N* = 462)		Male (*N* = 274)			
	M	*SD*	*M*	*SD*	*d*	*M*	*SD*	*M*	*SD*	*d*	Diff *d**
R	-0.29	0.86	0.49	1.02	-0.85	-0.27	0.91	0.45	0.98	-0.77	0.08
I	-0.27	0.90	0.45	0.99	-0.76	-0.23	0.92	0.38	1.01	-0.64	0.12
A	0.32	0.92	-0.53	0.89	0.93	0.29	0.96	-0.48	0.88	0.83	0.10
S	0.25	0.95	-0.42	0.94	0.70	0.25	0.95	-0.42	0.94	0.70	0.00
E	0.03	1.03	-0.05	0.94	0.08	0.03	1.04	-0.04	0.94	0.07	0.01
C	0.06	1.02	-0.10	0.96	0.16	0.06	1.03	-0.10	0.94	0.17	-0.01

Note. DIF = differential item functioning; GIST-R = Revised General Interest Structure Test; *M* = mean; RIASEC = Realistic, Investigative, Artistic, Social, Enterprising, and Conventional; *SD* = standard deviation.
Diff *d** represents the differences between the effect sizes (absolute values) in the standard test and the DIF-optimized test version.

Differential Validity

C-Indices as congruence indicators rest upon Holland's assumption of a circumplex model structure of vocational interests. In order to examine the extent to which the hypothesized order relation are held, randomization tests were implemented using the RANDALL program (T. J. Tracey, 1997). Hubert and Arabie (1987) proposed the correspondence index (CI) to assess model fit. The CI can range from -1.0 (all order predictions violated) to 1.0 (perfect model fit). In their meta-analytical study, T. J. Tracey and Rounds (1993) established CI values above .65 as a benchmark for good model fit. In our study, CI values ranged from .53 to .86 indicating sufficient model fit (see Table 10). The fit to the male data produced slightly lower CI values (CIs of .61 and .53) as was true for the female data (CIs of .86 and .75), as indicated in previous research (Darcy & Tracey, 2007).

C-Indices were calculated and then tested for significant gender differences to investigate Hypothesis 1. Results indicated significant differences between females (M = 11.38, SD = 3.32) and males (M = 11.94, SD = 3.99) in the standard version (GIST-R) with men showing a higher level of congruence, $t(733) = 2.04$, $p = .04$. Even though the magnitude of mean differences was more likely small (d = 0.15), no significant differences were found examining the DIF-optimized version (DIF-GIST-R).

Furthermore, analyses were repeated separately for each training course and field of study (see Table 11). Significant gender differences were identified in two training courses (digital media design [MD] and engineering drawing [ED]) and two fields of study (C and I) for the GIST-R with males showing higher levels of congruence than females. For the DIF-GIST-R, significant gender differences were found for two training courses (MD and ED) and one field of study (I). Thus, the DIF-optimized version led to a modest reduction of effect sizes.

Correlation coefficients between C-Index and the criteria satisfaction and subjective person-environment fit for both versions of the inventory are shown in Table 12. Correlations between C-Index and criteria are slightly higher for females than for males. Furthermore, correlations between C-Index and criteria for the standard version appear slightly lower than those for the DIF-optimized version of the instrument. Nevertheless, when testing the differences between the two correlations using Fisher r-to-z transformation neither comparison reached significance (z = 0.77 to 1.21; p = .11 to .22).

Table 10. *Correspondence Indices for Total, Female and Male Sample for the GIST-R and the DIF-GIST-R*

	GIST-R		DIF-GIST-R	
Sample	CI	*p*	CI	*p*
Total (N = 736)	.61	.017	.56	.017
Female (N = 462)	.86	.017	.75	.017
Male (N = 274)	.61	.017	.53	.017

Note. CI = correspondence index; DIF = differential item functioning; GIST-R = Revised General Interest Structure Test. CI (ratio of predictions met-predictions violated over total number of predictions).

Table 11. *Differential Effects Regarding the C-Index for the Standard Item-Set and the DIF-Optimized Item-Set*

Field of Study		*N*	C-Index GIST-R *M*	*SD*	*d*	*p*	C-Index DIF-GIST-R *M*	*SD*	*d*	*p*
MD	Female	71	11.63	3.79	0.77	.000	11.07	4.50	0.64	.002
	Male	39	8.56	4.13			8.36	4.00		
HM	Female	113	11.15	3.15	0.25	.245	10.87	3.18	0.13	.561
	Male	26	10.35	3.22			10.46	3.24		
ED	Female	55	11.40	3.82	-0.65	.000	11.44	3.50	-0.59	.001
	Male	70	13.80	3.59			13.50	3.45		
C	Female	47	12.28	3.48	-1.18	.011	13.11	3.39	-0.25	.184
	Male	67	13.91	3.24			13.93	3.10		
E	Female	72	10.96	3.20	-0.03	.890	10.71	2.99	0.02	.915
	Male	57	11.04	3.01			10.65	3.27		
L	Female	104	11.35	2.86	0.57	.024	11.29	3.05	0.62	.020
	Male	15	9.47	3.70			9.27	3.41		

Note. DIF = differential item functioning; GIST-R = Revised General Interest Structure Test; MD = Digital media design, HM = Hotel management, ED = Engineering drawing, C = Chemistry, E = Economics, L = Linguistics; *M* = mean; *SD* = standard deviation.

Table 12. *Differential Validity of the C-Index Based on the Standard and on the DIF-Optimized Item Set*

	1	2	3	4
1. C-Index GIST-R		.75**	.10	.14*
2. C-Index DIF-GIST-R	.69**		.15*	.19**
3. Satisfaction	.14**	.14**		.46**
4. Person-environment fit	.22**	.24**	.39**	

Note. DIF = differential item functioning; GIST-R = Revised General Interest Structure Test. Women are shown beneath the diagonal. Men are shown above the diagonal.
* Correlation is significant at the .05 level (two-tailed).
** Correlation is significant at the .01 level (two-tailed).

Regression Models

Control Variables. We found a significant effect of cognitive ability for design draftspeople (b = .04, p = .002, ΔR^2 = .073; see Table 13).

MMRs of the GIST-R. Step 2 of the MMRs showed a significant first-order effect of interest for digital media designers (b = .05, p = .013), design draftspeople (b = .05, p < .001) as well as for chemistry (b = .10, p < .001), economics (b = .04, p = .001), and linguistic students (b = .03, p = .014). In addition, cognitive ability was indicated as a significant predictor for design draftspeople (b = .03, p < .025). Overall, no significant effect of gender was found.

Step 3 shows the results after the interaction term was entered to the equation. Only for chemists, a marginal moderating effect (interaction term: b = .08, p = .053) was found in the standard test version. The interaction term accounts for an additional 2.7%, F(1, 108) = 3.84, p < .001, of the variance in person-environment fit. Thus, the relationship between interest and person-environment fit increase is stronger for females than for males. This result suggests the presence of differential prediction.

MMRs of the DIF-GIST-R. In the DIF-optimized version, significant first-order effects of interest are shown for design draftspeople (b = .10, p < .001) as well as for chemistry (b = .14, p < .001), economics (b = .06, p = .003), and linguistic students (b = .04, p = .005). In sum, the relationship between interest and the criterion was weaker than those found for the GIST-R.

Compared to the significant interaction in the subgroup of chemists in the GIST-R, the moderating effect decreases in the DIF-optimized version (b = .08, p = .093). This results in a reduction in ΔR^2. Now, 1.9% of the variance in person-environment fit is explained by the interaction term. Overall, no moderating effects of gender occurred in all other DIF-optimized versions (all ps > .05).

Table 13. *Results from the MMRs Separately for Each Field of Study Regarding the GIST-R Standard Version and DIF-Optimized Version*

	Model MD		Model HM		Model ED		Model C		Model E		Model L	
	b	ΔR^2	*b*	ΔR^2	*b*	ΔR^2	*b*	ΔR^2	*b*	ΔR^2	*b*	ΔR^2
GIST-R												
Step 1		**.007**		**.004**		**.073*****		**.030***		**.011**		**.007**
Cognitive Ability	.01		-.01		.04***		.04*		-.02		.01	
Step 2		**.062***		**.033**		**.111*****		**.171*****		**.084*****		**.059****
Cognitive Ability	.01		-.01		.03**		.04*		-.02		.01	
Interest	.05**		.02*		.05***		.10***		.04***		.03**	
Gender	-.03		-.25		-.10		.12		-.13		-.32	
Step 3		**.001**		**.000**		**.001*****		**.027*****		**.001****		**.006***
Cognitive Ability	.01		-.01		.03**		.04**		-.02		.01	
Interest	.04		.02		.04**		.06**		.05**		.06*	
Gender	-.01		-.25		-.09		.17		-.13		-.37	
Interest x Gender	.01		.00		.01		.08*		-.01		-.03	
DIF-GIST-R												
Step 1		**.007**		**.004**		**.073*****		**.030***		**.011**		**.007**
Cognitive Ability	.01		-.01		.04***		.04*		-.02		.01	
Step 2		**.035**		**.024**		**.120*****		**.248*****		**.072****		**.074****
Cognitive Ability	.01		-.01		.03**		.03		-.02		.01	
Interest	.03*		.02		.10***		.14***		.06***		.04***	
Gender	.05		-.26		-.14		.04		-.12		-.29	
Step 3		**.008**		**.000**		**.004*****		**.019*****		**.000****		**.013****
Cognitive Ability	.01		-.01		.03**		.04*		-.02		.01	
Interest	.01		.03		.08**		.10***		.06*		.10**	
Gender	.10		-.27		-.13		.06		-.12		-.31	
Interest x Gender	.04		-.01		.04		.08*		.00		-.06	

Note. DIF = differential item functioning; GIST-R = Revised General Interest Structure Test; MMR = multiple regression framework.
MD = Digital media design, HM = Hotel management, ED = Engineering drawing, C = Chemistry, E = Economics, L = Linguistics.
All quantitative predictors are mean-centered. ***** $p < .10$. ****** $p < .05$. ******* $p < .01$.

3.2.4 Discussion

Summary

Both NIE and APA Standards (NIE, 1975) strongly emphasize the suspension of gender differences in instrument validity and prediction. Nevertheless, we addressed the issue of differential validity and differential prediction jointly with a DIF analysis for the first time together in the vocational interest domain.

In order to demonstrate the absence of differential validity, the correlation between an individual's test result and the criterion must be equal for females and males. Thus, congruence indices indicating person-environment fit should be equal for females and males. Nevertheless, we established significant differences in the C-Indices for females and males when analyzing the total sample. Analyses on the group level found differential validity for media design and draftspeople trainees, as well as chemistry and linguistic students. Furthermore, we found a slightly higher correlation between C-Index and both criteria (subjective person-environment fit and satisfaction) for females than for males. Thus, evidence for differential validity was found. The overall mean correlation between congruence and satisfaction is comparable to meta-analytical findings (Tranberg et al., 1993; Tsabari et al., 2005).

Current research (Aros et al., 1998; Einarsdóttir & Rounds, 2009) on DIF suggests that women and men respond differently to certain interest items in traditional interest inventories albeit possessing the same underlying trait level. Eliminating those items showing large DIF led to a reduction of gender-specific mean differences in certain interest domains (namely, Realistic and Investigative interests). Furthermore, as Meade and Tonidandel (2010) highlight, removing items showing gender-specific DIF reduces measurement bias thereby eliminating one source of prediction bias. When analyzing the DIF-optimized version of the instrument, we found no significant gender differences in the C-Index for the total sample. Nevertheless, when assessing differential validity on the group-level gender differences in C-Indices remained significant for media design and draftspeople trainees as well as linguistic students. Interestingly, when comparing gender-specific correlation coefficients for C-Index and both criteria (subjective person-environment fit and satisfaction), we established slightly higher correlations for the DIF-optimized version than the standard version of the interest inventory. Thus, reducing measurement bias (i.e., removing those items showing large DIF) eliminated differential validity in the total sample and reduced differential validity on the group level.

We second focused on another source of prediction bias so far uninvestigated in the domain of vocational interest namely differential prediction. Differential prediction refers to differences in the regression lines (i.e., the slopes and intercepts of the regression model) for females and males. When examining the prediction of person-environment fit by an individual's interest score, we found a (marginal) moderating gender effect for the group of chemists. A significant interaction term indicated that the predictive relationship between interest and person-environment fit differed for females and males. When subsequently analyzing the DIF-optimized test version, no significant prediction bias was found. Thus, while there was some evidence for differential prediction in the standard version, prediction bias was eliminated in the DIF-optimized version.

Overall, we found evidence of differential validity and some indication for differential prediction in a standard Holland interest inventory. However, evidence for differential prediction was found only in one of the six groups. Future studies should examine whether this result can be replicated with other samples and other instruments. Nevertheless, we found evidence that removing those items showing large gender-specific DIF seems to be one strategy to diminish or even eliminate those prediction biases. Since testing for DIF is more of a straightforward procedure, it should be considered a standard in the interest domain.

Limitations

However, we are unable to determine whether our criteria–satisfaction and subjective person-environment fit–are themselves biased especially since they rely on self-reported measures. It could well be possible that those criteria are confounded with a variety of other constructs such as overall satisfaction with life and evaluation through significant others that are prone to stereotyping.

Furthermore, we are aware of relatively small sample sizes of females and males within each training course or field of study. Moreover, distributions between females and males within some groups are more likely disproportionate. This inhomogeneous distribution of gender might result in a lack of power of the MMRs (Aguinis, 2004). Furthermore, although different fields of vocational training and fields of study were investigated, the Social type seems to be slightly underrepresented in our sample. Thus, future research should first rely on more equally distributed and second in terms of dominant interest letter on more heterogeneous samples.

Implications

With respect to test construction principles, we recommend that interest inventories should by default be evaluated for gender-specific DIF using IRT methodologies since they directly assess the relationship between observed item responses and latent traits that are measured. Testing for DIF enables test developers to determine whether the test behaves differently for women and men and should be integrated into test analyses as a standard procedure. Second, as highlighted by Meade and Tonidandel (2010), both measurement bias and prediction bias must be inspected when examining test bias. Even though eliminating those items showing large DIF reduced prediction bias in the instrument, we also established differential validity in the DIF-optimized version on the group level. Thus, it remains unclear how the instrument can be further modified to guarantee test fairness. Further research concerning differential validity in interest inventories as well as its causes and possible moderators is needed.

Moreover, practitioners applying interest inventories should be aware that interest items, especially in the Realistic domain, are interpreted differently by female and male test takers partly explaining gender-specific mean differences on the RIASEC dimensions. In our study, removing those items showing large DIF led not only to a reduction in gender-specific mean differences in the Realistic, Investigative, and Artistic but also to a fairer prediction of the criteria person-environment fit and satisfaction. However, these results need to be replicated with an enlarged and more heterogeneous sample as well as with other Holland-based interest instruments.

3.3 Does gender-specific differential item functioning affect the structure in vocational interest inventories?

3.3.1 Aims and hypotheses

As the issue of test fairness has received great attention in the field of vocational interests (AERA, APA, & NCME, 1999; NIE, 1975), we contribute to recent research by gaining insight to the question whether or not structural properties of an interest inventory are affected when items showing DIF are eliminated in the test construction process. So far few studies focused on the problem of gender-specific DIF in interest measurements (Aros et al. 1998; Einarsdóttir & Rounds, 2009). Two studies found substantial DIF for the majority of interest items in the SII (Harmon et al., 1994); however, neither study analyzed the consequences of DIF-reduction on the psychometric structure of the investigated instrument. Therefore, we raise the following questions: Do the structural properties of the Holland-based interest inventory change when we consider DIF or can the model assumptions be maintained? Thus, the goal of this study is to examine whether a standard interest inventory that operationalizes the Holland model shows gender-specific DIF and moreover, whether eliminating DIF-items leads to alterations of the instruments psychometric structure. In order to evaluate structural models, research relies on two main approaches: exploratory or confirmatory concepts (Nagy et al., 2010). Lately, confirmatory methods have been widely implemented as they allow comparisons between alternative structural models. In the domain of vocational interests, confirmatory factor analyses (CFA; Jöreskog, 1969) and randomization tests of hypothesized order relations (RTOR; Hubert & Arabie, 1987) are recommended as the core analytic methods (Rounds, Tracey, & Hubert, 1992).

The study is based on the following publication:
Beinicke, A., Pässler, K., & Hell, B. (2014). Does gender-specific differential item functioning affect the structure in vocational interest inventories? *International Journal for Educational and Vocational Guidance*, *14*(2), 181-198. doi:10.1007/s10775-013-9254-y
The final publication is available at link.springer.com.

3.3.2 Methodology

Participants

The study involved a sample of 736 German students. Among these, 462 were women (62.77%) and 274 were men (37.23%). In order to guarantee sufficient heterogeneity of job-related interests, we investigated different fields of vocational training (digital media designers [15.1%], hotel managers [18.7%], and design draftspeople [17.0%]) and different fields of study (chemistry [15.5%], economics [17.5%], and linguistics [16.2%]). Participants were either enrolled in vocational training (vocational track: 50.8%) or university programs (academic track: 49.2%). Participants ranged from 16 to 48 years of age (M = 21.42, SD = 2.89). Data were collected online over a 6-month period.

Instrument

Revised General Interest Structure Test (GIST-R). Participants' occupational interests were assessed using the GIST-R (Bergmann & Eder, 2005). The GIST-R is a widely used German interest inventory based on Holland's RIASEC model (J. L. Holland, 1959, 1997). Sample items for the six interest types of the GIST-R are presented in Table 14. Participants are asked to rate their individual level of interest in the represented activity on a 5-point Likert scale, ranging from 1 ("I am not interested in this at all; I do not enjoy doing this at all.") to 5 ("I am very interested in this; I enjoy doing this very much."). Interest scores were aggregated to the six RIASEC dimensions (10 items per dimension). As indicated in the manual, all GIST-R scales have good reliabilities (1-month retest reliabilities range from r = .85 to r = .92). Correlations between matching scales with an adaptation of Holland's SDS instrument range from r = .60 to r = .75 (Jörin et al., 2003). Furthermore, positive validity information is provided in the manual. Studies support the hexagonal structure of the GIST-R for both female and male samples (Bergmann & Eder, 2005; Nagy et al., 2010).

Procedures and preliminary analyses

First, we investigated evidence for DIF in the GIST-R. Second, we analyzed the structural properties of the inventory across gender by applying different methodological approaches.

DIF-reduction or DIF-optimization. We carried out analyses of DIF in order to investigate whether GIST-R items assess the same underlying constructs for both women and men. We applied a non-parametric method for testing for DIF in polytomous items based on calculating the Liu-Agresti cumulative common log-odds ratio (L-A LOR; Liu & Agresti, 1996). This method is implemented in DIFAS 5.0 (Penfield, 2009) and is based on contingency tables. Derived from the Mantel-Haenszel common odds ratio used for dichotomous items, the L-A LOR is seen as its generalization for polytomous items (Penfield & Algina, 2006). In order to evaluate the DIF-size, we followed the classification system by the Educational Testing Service (Zieky, 1993). Three categories were found: A for items with negligible DIF (L-A LOR < .43), B for items with slight to moderate DIF (.43 ≤ L-A LOR < .64), and C for items with moderate to large DIF (L-A LOR ≥ .64). The method is described in detail by Wetzel and Hell (2013). Although the Educational Testing Service recommends eliminating both items showing slight to moderate (B) and large (C) DIF, we decided to eliminate only those items classified as C because eliminating the B items simultaneously would have left an insufficient number of items in the Realistic domain (see Table 14). This procedure led to a reduced number of total items (put in brackets) in all but one RIASEC dimension in the DIF-reduced inventory: Realistic (5), Investigative (7), Artistic (8), Social (10), Enterprising (6), and Conventional (8).

Confirmatory factor analyses. In the next step, we analyzed patterns of covariations to determine the underlying structural model. Holland's circumplex theory proposes equality constraints between adjacent, alternate, and opposite interest dimensions. Parametric methods, such as confirmatory factory analysis, are recommended to test these equality assumptions (e.g., Rounds et al., 1992). Browne (1992, 1995) proposed a circular stochastic process model with Fourier series correlation function to evaluate model-data fit within Holland's theoretical framework. We applied this method with the CIRCUM program (Browne, 1995) which uses structural equation modeling to evaluate four successively more constrained circumplex models using maximum likelihood estimates: (a) the quasi-circumplex (i.e., unconstrained) model, (b) the quasi-circumplex model with equal communality constraints (i.e., equal radii), (c) the circulant model with equal spacing constraints, and (d) the circulant model with both equal spacing and equal communality constraints. Each model is fitted to each subsample.

To evaluate the model-data fit, we examined several fit indices. First, the chi-square index is reported as an absolute measure of model-data fit. Since the overall chi-square statistic is more likely sensitive to sample size and model complexity we also focused on the following common fit indices: goodness-of-fit-index (GFI; Jöreskog & Sörbom, 1986) and the root mean square error of approximation (RMSEA). The GFI is an indicator of the variance accounted for by the model and ranges from 0 (poor fit) to 1 (perfect fit). The RMSEA is a badness-of-fit indicator. RMSEA values less than .05 indicate very close model-data fit, values between .05 and .08 good fit, values between .08 and .10 mediocre fit, and values above .10 poor fit (Browne & Cudeck, 1993; MacCallum, Browne, & Sugawara, 1996).

Randomization test of hypothesized order relationships. As proposed by Hubert and Arabie (1987), the randomization test of hypothesized order relations is an adequate method to test order relations within a RIASEC correlation matrix (Wakefield & Doughtie, 1973). The test first provides the probability of obtaining the predicted order and second a fit index (Correspondence Index, CI) that measures how the collected data correspond to the hypothesized model (Rounds et al., 1992). A CI = -1 means that no order prediction was met; a CI = +1 shows a total match of order predictions. In their meta-analysis Rounds and Tracey (1996) proposed benchmarks for a good model fit to the Holland model: For U.S. samples, a CI of .70 and for international samples, a CI of .48 were found. Correlation matrices of the RIASEC dimensions provided the basis of our analysis. The randomization test was applied separately for the total sample and each gender-specific subsample testing Holland's circumplex model. Holland's theory assumes that the correlations for the six adjacent pairs (R-I, I-A, A-S, S-E, E-C, and C-R) should be greater than the correlations of the six alternate pairs (R-A, I-S, A-E, S-C, E-R, and C-I), and in turn those should be greater than those for the opposite dimensions (R-S, I-E, A-C). In total, the circumplex model provides 72 order predictions. We used RANDALL (T. J. Tracey, 1997) to conduct randomization tests and to test the equivalence of models across subsamples.

Table 14. *DIF Items GIST-R*

Scale	DIFAS DIF L-A LOR (SE)	A B C	Items in DIF-reduced inventory	Item content (A items only)
Realistic	0.69 (.32)	2 3 5	5	to work on a construction site investigate how something works
Investigative	0.22 (.11)	6 1 3	7	observe and analyze something meticulously investigate the causes of a problem
Artistic	0.29 (.14)	7 1 2	8	design something artistically do things that depend on creativity and fantasy
Social	0.08 (.05)	8 2 0	10	attend to, serve other people listen to other people's problems
Enterprising	0.18 (.09)	6 0 4	6	supervise or monitor others oversee a group at work
Conventional	0.23 (.11)	7 1 2	8	monitor the adherence to principles work in accounting

Note. Classification for L-A LOR: < .43 = A (negligible); .43 ≤ L-A LOR < .64 = B (slight to moderate); ≥ .64 = C (moderate to large). Positive values indicate DIF in favor of the reference group (men) and negative values indicate DIF in favor of the focal group (women). Item content consists of the GIST-R items translated into English.

Table 15. *Descriptive Statistics and Correlations for the Interest Dimensions*

	Females N = 462		Males N = 274			Correlations					
	M	*SD*	*M*	*SD*	*d*	R	I	A	S	E	C
GIST-R											
R	2.37	0.64	2.95	0.76	-0.83	-	.48**	.01	-.00	-.03	.22**
I	2.68	0.68	3.22	0.74	-0.76	.54**	-	.12	.13*	.12*	.16**
A	3.41	0.75	2.72	0.72	0.93	.00	.18**	-	.39**	.26**	.15*
S	3.40	0.71	2.91	0.70	0.70	.02	.07	.28**	-	.55**	.34**
E	3.31	0.78	3.25	0.72	0.08	.00	-.02	.10*	.54**	-	.48**
C	2.91	0.69	2.81	0.65	0.16	.18**	.16**	-.07	.38**	.55**	-
GIST-R DIF											
R	2.59	0.68	3.13	0.73	-0.76	-	.40**	-.06	.06	.04	.22**
I	2.88	0.80	3.41	0.87	-0.63	.47**	-	.09	.14*	.17**	.11
A	3.23	0.78	2.60	0.72	0.84	-.08	.12**	-	.40**	.31**	.11
S	3.40	0.71	2.91	0.70	0.70	.05	.12*	.29**	-	.55**	.33**
E	3.25	0.84	3.20	0.76	0.07	.07	.09*	.13**	.53**	-	.43**
C	2.94	0.72	2.82	0.65	0.17	.17**	.14**	-.06	.39**	.52**	-

Note. RIASEC refers to the 6 vocational interest types as follows: R = Realistic; I = Investigative; A = Artistic; S = Social; E = Enterprising; C = Conventional. For females correlations are shown below the diagonal and for males above the diagonal. Females and males differ regarding all RIASEC-dimensions (all *p*s < .05); only in the Enterprising domain no differences were found across gender.

3.3.3 Results

Descriptive statistics for the GIST-R revealed gender differences in RIASEC dimensions that are comparable to those consistently found in vocational interests inventories (Lippa, 1998; Su et al., 2009). Particularly, men scored higher on Realistic and Investigative dimensions, whereas women scored higher on Artistic and Social dimensions. These gender differences were slightly reduced in some dimensions (e.g., R, I, and A), when items showing large DIF are eliminated (see Table 15). Our findings are in line with results found by Aros et al. (1998) and Einarsdóttir and Rounds (2009). Additionally, for women and men all mean differences between the standard and the DIF-optimized test are small to negligible (Cohen's $d \leq 0.33$). When analyzing the correlational patterns of the interest dimensions we found roughly the assumed circular order for both females and males; e.g., adjacent scales correlated most highly, whereas correlation decreased between remote scales. Merely, the correlation between I and A was more likely small in size as already noticed by Nagy et al. (2010).

Confirmatory factor analysis

Results of the confirmatory factor analyses are presented in Table 16. As recommended by Fabrigar, Visser, and Browne (1997), we tried different specifications for the "m" parameter. We examined each of the four different models with m equal to 1, 2, and 3. Since values of m = 2 resulted in best model-data fit, these analyses are presented in Table 16. When examining the equal communality quasi-circumplex model we found a non-significant chi-square ($\chi^2 = 20.16$, $df = 7$, $p = .44$) for the overall sample. Furthermore, both the GFI of .99 and the RMSEA of .05 indicated a good fit to the data. For the male sample the equal communality quasi-circumplex model also showed a good model-data fit ($\chi^2 = 31.63$, $df = 7$, $p = .09$; GFI = .99; RMSEA = .07). For the female data we found evidence for a good model-data fit only for the less constrained quasi-circumplex model ($\chi^2 = 13.83$, $df = 2$, $p = .05$; GFI = .99; RMSEA = .09). Thus, males appeared to be slightly better described by the equal communality quasi-circumplex model than females. Results indicated poor model-data fit for the more constrained equal spacing or circulant model for all samples (all chi-square statistics were significant, GFI indices were below .96, and all RMSEAs were well above the threshold of .05).

Table 16. *Summary of Fit Indices for the Confirmatory Factor Analyses of Holland's Hexagonal Model*

		CIRCUM											
		Quasi-circumplex model			Quasi-circumplex model with equal communalities			Circulant model (equal spacing)			Circulant model with equal communalities		
	n	GFI	RMSEA	$\chi^2(2)$	GFI	RMSEA	$\chi^2(7)$	GFI	RMSEA	$\chi^2(7)$	GFI	RMSEA	$\chi^2(12)$
GIST-R													
Female	462	.99	.09	13.83*	.99	.08	38.62	.94	.16	144.51	.91	.16	229.22
Male	274	.99	.09	14.75	.99	.07	31.63*	.95	.15	117.36	.92	.14	195.75
Total	736	.99	.14	30.70	.99	.05	20.16*	.90	.21	242.18	.88	.19	326.25
GIST-R DIF													
Female	462	.99	.14	32.54	.89	.23	278.27	.94	.16	140.72	.91	.16	230.48
Male	274	.99	.10	15.60	.99	.07	29.73*	.95	.15	120.27	.93	.14	189.53
Total	736	.99	.09	14.35*	.93	.18	180.29	.91	.21	229.17	.88	.18	304.84

Note. 90% confidence intervals for the root-mean-square error of approximation (RMSEA) have been omitted to save room. GFI = goodness-of-fit-index. * $p > .05$.

Comparison between standard and DIF-reduced test version. In the DIF-reduced test version we found a comparable model-data fit for the quasi-circumplex model with equal communalities in the male sample (χ^2 = 29.73, *df* = 7, *p* = .12; GFI = .99; RMSEA = .07). For the total sample only the unconstrained quasi-circumplex model now seemed to fit the data well (χ^2 = 14.35, *df* = 2, *p* = .05; GFI = .99; RMSEA = .09). However, neither model indicated good model-data fit for the female subsample in the DIF-reduced test version (all chi-square statistics were significant, GFI indices between .89 and .99, and RMSEAs between .14 and .23).

To further investigate the differences between the standard and the DIF-optimized version we examined the angular estimates provided by the quasi-circumplex model and the quasi-circumplex model with equal communalities (see Table 17). Within an equal spacing model interest dimensions are assumed to be 60° apart. For the equal communality model the total sample demonstrated smaller gaps between R and I, between S and E, and between E and C than the expected 60°. Moreover, there was a large gap between I and A as already indicated by the analysis of the correlation matrix. This pattern also appeared for the male sample (equal communality model) and the female sample (unconstrained model). A similar pattern applied for the male sample in the DIF-optimized interest version. Besides, for both the female and the total sample we found order violations of the interest dimensions. These deviations might account for the poor model-data fit in these samples. Moreover, males and females differed in the severity of these deviations from an equally spaced circle.

For the more constrained circulant models fit indices indicated poor model-data fit. Therefore, we did not further analyze these two models. Since investigations of the RIASEC structure have generally assessed the fit of the circulant model (Darcy & Tracey, 2007) our results are inconclusive. Although results of the confirmatory factor analyses indicated that DIF-reduction did not decline the overall model fit, Holland's strict structural assumptions are not supported by the data. Moreover, analysis of the angular estimates showed severe deviations from the assumed equal spacing of the interest dimensions in the female sample.

Randomization test of hypothesized order relationships

The results of the randomization test are summarized in Table 18. Overall, the test indicated a good fit for the hexagonal model at the level of the total sample as well as for each of the subsamples (women and men) in both test versions (*p* = .02). This probability value suggested that the number of predicted order relations violated was significantly smaller than the number that would be obtained under the random labeling hypothesis. Thus, the null hypothesis of random labeling was rejected for the entire sample as well as for the subsamples. The CIs for Holland's model for the male and the total sample were lower than the U.S. benchmark values (CI = .70) provided by Rounds and Tracey (1996) but higher than their benchmark for an international sample (CI = .48). The fit of the model to the female data produced slightly higher CI values ($CI_{standard ♀}$ = .86) than for the male data ($CI_{standard ♂}$ = .61*)*. To examine the difference in fit across gender, we conducted a randomization test of differences in fit. For these analyses corresponding pairs of correlation matrices are considered together to test differences in fit and *p*-values for each comparison provide the probability that any difference in fit was obtained by chance (for details, see Anderson et al., 1997). The results indicated no significant differences in model fit between women and men (CI = .12, *p* = .10).

Comparison between standard and DIF-reduced test version. For the total sample, analyses of the correspondence indices across test versions (standard vs. DIF-reduced) revealed comparable results in model-data fit; no significant differences were found, all *p*s > .10. Neither for the female, nor for the male or the total sample, the differences in the CIs were significant. In women, we found a slight but non-significant decline in the CIs in the DIF-optimized version ($CI_{standard ♀}$ = .86 and $CI_{DIF ♀}$ = .75, Δ = .11; CI = .06, *p* = .18). Likewise for the male data, we found a slight but again non-significant decline in the CIs ($CI_{standard ♂}$ = .61 and $CI_{DIF ♂}$ = .53, Δ = .08; CI = .04, *p* = .22). In sum, randomization tests indicated that for both females and males the DIF reduction did not change the hypothesized correlation order in a statistically significant degree.

Table 17. *Angular Placement (in Degrees) of Interest Types from CIRCUM Analysis*

	CIRCUM											
	Quasi-circumplex model						Quasi-circumplex model with equal communalities					
	R	I	A	S	E	C	R	I	A	S	E	C
GIST-R												
Female	0	19	123	204	226	273	0	20	121	200	229	265
Male	0	49	164	185	214	271	0	40	155	212	236	277
Total	0	354	190	223	253	280	0	12	150	206	240	267
GIST-R DIF												
Female	0	341	145	107	85	61	0	191	304	86	259	238
Male	0	52	212	218	236	285	0	49	164	214	231	277
Total	0	351	192	245	267	284	0	190	264	283	314	251

Note. RIASEC refers to the 6 vocational interest types as follows: R = Realistic; I = Investigative; A = Artistic; S = Social; E = Enterprising; C = Conventional.

Table 18. *Summary of Fit Indices of the Randomization Test of Holland's Hexagonal Model*

	Randomization Test (RANDALL)			
	n	Ratio	CI	*p*
GIST-R				
Female	462	67-0	.86	.0167
Male	274	58-0	.61	.0167
Total	736	58-0	.61	.0167
GIST-R DIF				
Female	462	63-0	.75	.0167
Male	274	55-0	.53	.0167
Total	736	56-0	.56	.0167

Note. CI = correspondence index (ratio of predictions met - predictions violated over total number of predictions).

3.3.4 Discussion

Main findings

The present study investigated whether the structural assumptions of a typical interest inventory are maintained when items showing large gender-specific DIF are eliminated. We used CFAs and randomization tests to establish the structural validity of an interest measure and its DIF-optimized version. When assessing Holland's structural assumptions randomization tests indicated a good fit of the data for both the total sample as well as for each of the subsamples (women and men); no significant differences in model fit between women and men were revealed. Results of the parametric CIRCUM analyses vary depending on the level of restrictions. When approaching the less constrained quasi-circumplex models, the model-data fit was adequate or good. However, the results were less supportive for the more constraint circulant models.

Past research provided evidence that different methodological approaches can lead to different assumptions regarding the structural representation of vocational interests (Darcy & Tracey, 2007; Gupta, Tracey, & Gore, 2008; Nagy et al., 2010). Considerable differences are found between parametric (e.g., CFA) and non-parametric (e.g., randomization test) methods. As Nagy et al. (2010) highlighted, whereas randomization tests are a non-parametric approach evaluating order predictions by analyzing the correlations between interest domains, parametric methods evaluate a response model that links latent variables and observed responses. Furthermore, Darcy and Tracey (2007) pointed out that both methods also address different research questions. Non-parametric tests compare the fit of the model to the fit attainable by chance. Parametric approaches assess model fit compared to perfect model-data fit. Given these methodological differences it might be reasonable to assume that there are differences in results.

Nevertheless, studies investigating the structural properties of the GIST-R generally supported Holland's hexagonal model for different age groups and provided further evidence for structural gender invariance (Bergmann & Eder, 2005; Nagy, 2006; Nagy et al., 2010). Furthermore, in Germany the structural validity of the Holland model has been demonstrated with other instruments besides the GIST-R (e.g., Jörin Fux, 2005). Whereas Nagy et al. (2010) investigated a large heterogeneous sample, we examined a more likely homogeneous sample that was preselected with regard to specific training courses and fields of study. Thus, problems in finding support for the constrained equal spacing model with the CFAs might be partly due to our sample characteristics. Moreover, as highlighted by Darcy and Tracey (2007) non-parametric approaches might be more suitable to assess model fit within the Holland framework since even slight deviations from the perfect model-data fit will lead to poor goodness-of-fit measures in the CFAs. Since, first, the Holland model was supported with German samples and for the instrument measure applied (see also Nagy et al., 2010) and second, the RTOR analyses also supported the hexagonal model, we concluded that the poor fit of the constrained circulant models in the CFAs might be partly attributable to our sample characteristics and methodological differences.

Our results regarding the structural invariance when eliminating items showing large DIF varied upon the method applied. The non-parametric randomization test indicated a good model-data fit. All CIs for both the standard and the DIF-optimized test version were above the benchmark for international samples (CI = .48) defined by Rounds and Tracey (1996) as an indicator for a good model fit but lower than the U.S. benchmark (CI = .70). Nevertheless, we found evidence for gender invariance for Holland's circumplex model regarding both the standard and the DIF-optimized test version. Our results using randomization tests indicated that the DIF-reduction did not affect the model's structure. Furthermore, randomization tests also demonstrated that gender-specific structural invariance can be assumed for both the standard instrument and its DIF-optimized version.

However, the results of the CIRCUM analyses were less supportive for the fit of Holland's model to the data. Neither of the circulant models (equal spacing and with equal communalities) showed a good fit for the GIST-R DIF. Moreover, analyses of the angular estimates revealed considerable deviation from the spacing assumption yielded by a circulant model, for example, each RIASEC dimension is separated by 60°. Whereas R and I as well as S and E, and E and C scales were less distinct than expected, we found a large gap between I and A. Also, neither the male nor the female data fitted the more constrained equally spaced models. Consequently, we were unable to test for gender invariance of the circulant models. CIRCUM analyses revealed that for both, the total and the female sample, the ordering of the letters violates Holland's structural assumption (e.g., RCESAI instead of RIASEC) in the DIF-optimized version. Moreover, for both the male and the female sample we found a closer than assumed proximity between R and I; thus, the usual distinctions between RIASEC dimensions is not as salient as Holland's model assumes. This close proximity between adjacent interest dimensions described by Armstrong et al. (2003) as *type compression* was also found in other studies (Hansen et al., 1993; Nagy et al., 2010). Whereas Nagy et al. (2010) showed this proximity also investigating the GIST-R, Armstrong et al. (2003) revealed type compression between R and I investigating a variety of samples completing either the SII (Harmon et al., 1994) or the UNIACT-R (ACT Program, 1995). Generally, there is a variety of studies that found a so-called *misshapen polygon* when analyzing the underlying structure of Holland based interest inventories (also see J. L. Holland, 1997; J. L. Holland & Gottfredson, 1992; Rounds & Day, 1999). As mentioned before, parametric analyses test for a perfect model-data fit; thus, these constrained model assumptions might be too restrictive for our preselected sample. Nevertheless, although the strict assumptions of the circulant models cannot be met by our data; we found no decline in the fit indices when comparing the standard and the DIF-optimized instrument version.

Considering our research question we conclude that both methods give evidence that when items showing large DIF are excluded from the interest measure the psychometric properties of the instrument are maintained. Thus, our results show that controlling for DIF is one possibility to reduce gender biases in interest inventories as claimed by test development standards (AERA, APA, & NCME, 1999; NIE, 1975) while leaving structural properties of the instrument intact. Furthermore, there is evidence that compared to other approaches to achieve gender fairness in interest inventories, such as balancing scales, controlling for DIF proved to be the most useful strategy to optimize gender fairness in the test development process (Wetzel et al., 2012) and will not impact the instrument's predictive validity (Pässler et al., 2014). Implications for practice are discussed below.

Limitations and future research

In the following, we address factors that limit our conclusions and suggest some ideas for future investigations and how they might cover these limitations. We investigated a sample comprising women and men that were in equal shares either enrolled in different fields of vocational training or in different fields of university programs. Analyzing this sample, we tried to guarantee sufficient heterogeneity of job-related interests; however, a bigger sample with a larger variety of target groups that is not limited to a German population is suggested in further investigations in order to generalize our results. Methodologically, we conducted DIF-reduction based on the recommendations of the Educational Testing Service (Zieky, 1993). Due to an insufficient minimum number of items in the Realistic domain, we eliminated only items of the category C (moderate to large DIF; L-A LOR ≥ .64). As suggested from the Educational Testing Service, future research studies–if possible–should additionally eliminate B-items (slight to moderate DIF; .43 ≤ L-A LOR < .64). In our study, we focused on the most cited and well-known interest model. However, future research is needed to understand how other interest models, such as Gati's (1979) or Rounds and Tracey's hierarchical model (1996), respond to DIF-reduction.

We also limit our conclusions to the GIST-R (Bergmann & Eder, 2005); another aspect that needs further exploration is the systematic comparison of the susceptibility of different item formats for DIF. The way the items are presented to the user might affect the results in terms of structural coherence: Items referring to behavior descriptions (e.g., GIST-R; Bergmann & Eder, 2005; EXPLORIX; Jörin et al., 2003) differ from items that present job titles (SII; Harmon et al., 1994). Einarsdóttir and Rounds (2009) found evidence that gender differences are reduced when items showing large DIF are removed especially in the Realistic dimension. Our results confirm these findings that gender-DIF occurs on different item types, including descriptions of activities and occupations, and thus are not limited to occupational items. Consequently, in order to construct unbiased test items, it is necessary to standardly eliminate such items showing gender-specific DIF in vocational interest inventories.

As evidence in support of Holland's structural conceptualization varies considerably across cultures (e.g., Long & Tracey, 2006; Rounds & Tracey, 1996), DIF-reduction and its consequences should be further investigated cross-culturally. Considering both–gender and cultural variables–simultaneously should further increase our understanding of how structural properties of interest measures might be influenced by eliminating items showing large gender-specific DIF.

Implications for practice

By investigating whether the structure of vocational interest inventories is affected by gender-specific DIF, we refer directly to the issue of test fairness (AERA, APA, & NCME, 1999; NIE, 1975). In order to assert test fairness, it is necessary to show that the underlying conceptualizations including its assumptions are not violated and also do not differ across gender. When items showing large gender-specific mean differences are eliminated through the test construction process, the content range of the dimension is often considerably reduced. Eliminating items showing large DIF might be a good alternative to achieve gender fairness in interest inventories since items that are endorsed differently by women and men albeit having the same underlying trait level are excluded from the instrument, thereby decreasing one source of measurement bias as Meade and Tonidandel (2010) highlighted. Moreover, DIF-analyses should be complemented by the analysis of a possible predictive bias (AERA, APA, & NCME, 1999; Pässler et al., 2014). Taken together, these analyses should ensure fair interest assessment that is adequate for counseling purposes.

LATER CAREER STAGES: FROM TRAINING EVALUATION TO CAREER DEVELOPMENT

According to the rationale of career construction theory, it is of great interest *how* individuals maintain successful and satisfying work lives by adapting to the changing nature of work environment in the 21st century (Savickas, 2006). In later career stages, individuals must strive for career adaptability to best respond to rapidly changing labor needs of post-modern societies once an occupation is chosen to maintain career success. Therefore, individuals focus on active career development. A crucial building block of career development is acquiring new knowledge, skills, or abilities through the successful completion of trainings. In this regard, guaranteeing training success is important to ensure that individuals make satisfactory progress towards their career development. But when is it reasonable to say that training is successful? How is training success measured? What are possible predictors of training success? These questions lead directly to the issue of training evaluation which is the focus of interest in the following.

Chapter 4 begins with an integrative literature review of training evaluation. First, theoretical models of training evaluation are briefly described (section 4.1). Second, a review in light of more recent empirical studies of the determinants of training success with its facets success of learning and success of transfer is presented (section 4.2). In this section, the role of individual influences (e.g., trainee characteristics) and organizational influences (e.g., training design and work environment) are displayed within the current research literature. Third, evaluating training settings in practice with regard to their effectiveness, especially the shift from traditional classroom training to e-learning, is focused (section 4.3).

Chapter 5 provides empirical evidence of three studies (Studies 4, 5, 6) on training evaluation. The fourth study investigated separately for e-learning and classroom training the extent to what evidence-based actions for maximizing training effectiveness are already implemented before, during, and after training in companies that have been awarded as “best employers in Germany” (section 5.1). The fifth study explored differences in training success of trainees in a specific corporate e-learning and classroom training using a field experiment with a time-lag design (section 5.2). The sixth study empirically examined determinants that ensure training success in corporate e-learning (section 5.3).

CHAPTER 4 LITERATURE REVIEW OF TRAINING EVALUATION

4.1 Training evaluation theories

The purpose of this section 4.1 is to provide an overview of theoretical models of training evaluation research. As an introduction, the two major functions of training evaluation are briefly outlined. Then, three theoretical models describing the determinants and outcome measures of training success are delineated in closer detail: namely, (a) the four-level evaluation model by Kirkpatrick (1959) (section 4.1.1) that was later criticized by Holton (1996) who proposed (b) the HRD evaluation and research model (section 4.1.2), and (c) the transfer process model by Baldwin and Ford (1988) that encouraged research on training transfer (section 4.1.3).

Functions of training evaluation

As first suggested by Scriven (1967), evaluations are distinguished into two major functions of evaluation: *formative evaluation* and *summative evaluation*. The function of *formative evaluation* is to monitor how well training objectives are met while the training program is in progress. By identifying shortcomings, formative evaluation aims at optimizing the training program continually in a process-accompanying format. Formative evaluation focuses on analyses of target, environment, concept, and teaching process. For example, if training results or suggestions for improvements that are typically collected qualitatively are reported back to trainers or T&D professionals, the training program itself can enhance in form and content while being in progress (Solga, 2011b).

In contrast, *summative evaluation* is used to assess how well training results measured at the end of a training program meet formerly set objectives. Summative evaluation aims at measuring and contrasting outcomes. Summative evaluation focuses on analyses of effectiveness and efficiency. For example, data, typically collected quantitatively and according to research methodology standards, can be used for strategic decisions such as whether a training program should be continued or discontinued (Solga, 2011b). Summative evaluation can be classified into short-term outcome evaluation (e.g., reactions or learning outcomes of trainees) and long-term outcome evaluation (e.g., behavior on the job or organizational results) [G. G. Wang & Wilcox, 2006]. This leads directly to Kirkpatrick's model.

4.1.1 Four-level evaluation model by Kirkpatrick

One of the most recognized and early-proposed evaluation models for measuring the effectiveness of a training program is Donald Kirkpatrick's four-level evaluation model (Kirkpatrick, 1959, 1967, 1975, 1994). Despite the former popularity of his model, Kirkpatrick's ideas become later known to a broad audience when he published his book entitled "Evaluating Training Programs" in 1994. The model distinguishes between four different levels or steps that are hierarchically and causally sequenced. This sequence implies that success at a higher level is only possible if targets of the underlying level are successfully achieved (see Figure 3).

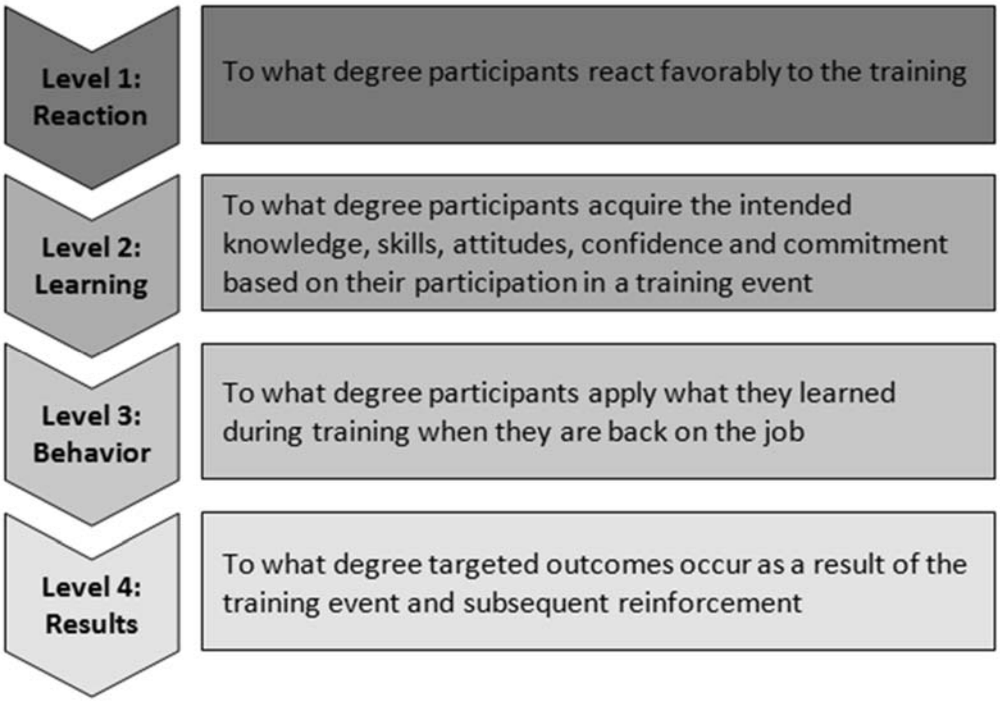

Figure 3. The Kirkpatrick model. Adapted from "The Kirkpatrick Model" by D. Kirkpatrick, 1954. Retrieved from http://www.kirkpatrickpartners.com/OurPhilosophy/TheKirkpatrick Model/tabid/302/ Default.aspx. Copyright 2009-2015 by Kirkpatrick Partners, LLC. Adapted with permission.

Level 1: Reaction–How well did trainees like the learning process?

This level investigates how trainees perceived the training. For example, trainees rate the correspondence between training content and individual expectations or the fit of the learning and working environment (Kirkpatrick, 1994). According to Alliger, Tannenbaum, Bennett, Traver, and Shotland (1997), the level reaction is further divided into utility and affective reactions such as whether training is perceived as enjoyable or stressful. Since trainees' reactions are available easily by using standardized questionnaires or oral feedback, these are the most commonly but sometimes the only measured evaluation criteria (Twitchell, Holton, & Trott, 2000). The more satisfied trainees are, the more effective is the training (Kirkpatrick, 1994). However, this hierarchical-causal dependency was refuted in a meta-analysis (Alliger et al., 1997). Here, empirical evidence revealed that affective reactions did not correlate with neither measures of learning (Level 2; r = .02) nor transfer of learning (Level 3; r = .07). This implies that affective reactions do not qualify for predicting outcomes of learning, transfer of learning, or corporate results (Tharenou, Saks, & Moore, 2007). According to the meta-analysis by Alliger et al. (1997), correlations were stronger regarding utility-type reaction measures and measures of either learning (Level 2; r = .26) or transfer of learning (Level 3; r = .18).

Level 2: Learning–What did trainees learn?

This level examines the extent to what trainees gain knowledge, skills, and attitudes. Here, Kraiger, Ford, and Salas (1993) proposed three categories of learning outcomes: cognitive, skill-based, and affective. The cognitive category subsumes verbal knowledge, knowledge organization, and cognitive strategies. Skill-based outcomes comprise skill compilation and automaticity. Finally, affective outcomes include attitudinal and motivational constructs such as disposition, self-efficacy, and goal-setting. Learning outcomes of training are usually measured using work samples, simulations or situational interviews, or performance tests that are preferably pre-post measures. Again, only weak to moderate relationships between learning and transfer of learning were found with correlations ranging from r = .08 to r = .18 (Alliger et al., 1997). This suggests that success of learning is a necessary but not sufficient condition for success of transfer (Solga, 2011b).

Level 3: Behavior–What changes in job performance resulted from the learning process?

This level measures the extent and quality to what trainees apply the newly acquired knowledge, skills, and attitudes to their daily work. In practice, T&D professionals refer to this level as transfer of learning which is of prime importance from an entrepreneurial point of view (Mandl, Prenzel, & Gräsel, 1992). Most behavioral data are collected using behavioral observations in the workplace either from supervisors, colleagues, or the trainees themselves. Weak to moderate correlations between learning and transfer of learning as mentioned above indicate that other constructs determine the success of transfer. Such determinants are subsumed in the transfer climate (J. B. Tracey, Tannenbaum, & Kavanagh, 1995; Solga, 2011b).

Level 4: Results–What are economic results?

The final level of Kirkpatrick's model examines the extent to what targeted outcomes result from of the learning process. For example, economic outcomes (e.g., individual or team productivity, quality, costs, work or customer satisfaction) are measured. Such visible outcome measures are often obtained later in performance reviews, economic analyses of operating numbers, or questionnaires investigating organizational climate (Solga, 2011b). However, it seems difficult to measure especially economic hard facts (e.g., reduced costs, turnover, improved quality, or increased production) because such indicators are often influenced by other extraneous factors that are not related to the training itself but rather depend on the cyclical economic trend or number of competitors which can lead to confounding effects (Goldstein & Ford, 2002).

Although Kirkpatrick's (1959) model of training evaluation is "elegant in its simplicity and has contributed greatly to HRD" (Holton, 1996, p. 6) and thus stimulated many important research studies, it remains flawed as an evaluation model according to Holton's criticism (Holton, 1996). Primarily, the model fails to meet criteria of good theories or models (for such criteria, see Klimoski, 1991, pp. 254-256). Here, one issue refers to the causal and hierarchical relationships between the levels. In Kirkpatrick's (1959) model, however, the implied causal linkages between the levels have not been demonstrated in empirical studies (Alliger & Janak, 1989; Holton, 1996). Moreover, as stated previously, correlations reported in empirical studies varied widely; thus, linear causal relationships seem doubtful and very unlikely. Since a theory implies empirical testing with correlational and experimental studies, Kirkpatrick's (1959) levels should rather be labeled as taxonomies implying only classification schemes instead of levels of a theory (Holton, 1996). Holton (1996) claims that a more complete model should move from this taxonomic evaluation approach to a fully specified model for HRD evaluation. Elaborating on the offered criticism, he finally proposed an integrative evaluation model that is presented next.

4.1.2 HRD evaluation and research model by Holton

Holton's (1996) HRD evaluation and research model is theoretically derived and conceptually comprehensive. The model integrates findings from empirical research studies with a grounded theoretical framework for diagnosing and understanding causal relationships between training outcomes. Three sequentially connected outcome levels are crucial: learning, individual performance, and organizational performance.

On a primary level, conceptual constructs in the domains of motivation, (work) environment, and ability influence each of the three training outcome levels directly but also indirectly via correlations with other primary constructs. On a secondary level, secondary influences, which are known as trainee characteristics, affect motivation and then further individual performance. Generally, this evaluation model was empirically supported. However, due to accumulating research evidence, the model was revised and modified by delineating specific constructs within the conceptual constructs. The following explanations of the model's structure are based on the revised HRD evaluation and research model (see Figure 4).

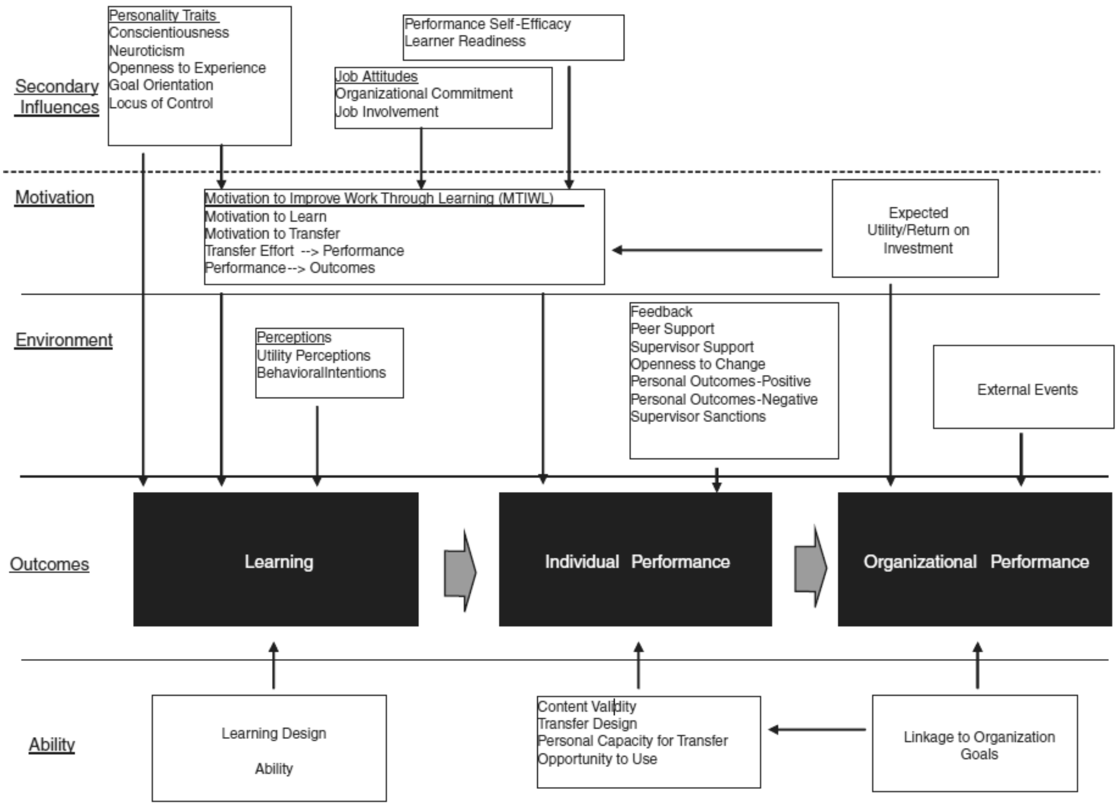

Figure 4. Revised HRD evaluation and research model by Holton (2005). Adapted from "Holton's evaluation model: New evidence and construct elaborations" by E. F. Holton, 2005, *Advances in Developing Human Resources*, *7*(1), p. 51. Copyright 2005 by Sage Publications. Reprinted with permission.

First, attention is drawn on *learning outcomes*. Since research focuses on trainees' dispositional characteristics referred to as individual stable traits of trainees, they are included in the model as secondary influences that affect learning through motivation to learn. Of the Big Five personality trait factors–(a) conscientiousness (Colquitt, LePine, & Noe, 2000), (b) neuroticism or emotional stability (Herold, Davis, Fedor, & Parsons, 2002), (c) openness to experience (Herold et al., 2002), (d) goal orientation (Bell & Kozlowski, 2002), and (e) locus of control (Colquitt et al., 2000)–only the latter two have received compelling empirical support in the research literature as measures of individual characteristics. In addition to personality traits, job attitudes (e.g., organizational commitment and job involvement) were identified as strong predictors of motivation to learn and motivation to transfer. Both are combined in the revised construct labeled motivation to improve work through learning. Referring to the level reaction in Kirkpatrick's (1959) model, trainees' perceptions of the training in terms of utility perceptions and behavioral intentions were found to significantly predict learning and performance outcomes and thus are included in the model (Tan, Hall, & Boyce, 2003).

Second, *individual performance outcomes* are primarily affected by transfer climate that was broadened to the learning transfer system and subsumes all factors in the person, training, and organization (Holton, 2005). Holton and his colleagues developed an instrument called Learning Transfer System Inventory (LTSI; Holton, Bates, & Ruona, 2000) measuring 16 constructs that are identified as critical barriers and catalysts of transfer (Bates, Kauffeld, & Holton, 2007; Wirth, Kauffeld, Bates, & Holton, 2009; see Figure 5). The LTSI is a standardized psychometric inventory and received strong empirical evidence of construct validity (e.g., Holton et al., 2000; Holton, Bates, Bookter, & Yamkovenko, 2007), criterion validity (e.g., Ruona, Leimbach, Holton, & Bates, 2002), and cross-cultural validity (e.g., Chen, Holton, & Bates, 2005 [Taiwan version]; Devos, Dumay, Bonami, Bates, & Holton, 2007 [French version]; Kauffeld, Bates, Holton, & Müller, 2008 [German version]; Khasawneh, Bates, & Holton, 2006 [Arabic version]; Yamkovenko, Holton, & Bates, 2007 [Ukrainian version]). The LTSI is based on the HRD evaluation and research model (Holton, 1996, 2005).

Third, *organizational performance outcomes* and their influences are rather poorly investigated. Even though the focus in training research is on the individual level, there is evidence that expected utility, return on investment (ROI), and a linkage to organizational goals influence organizational results (Bates & Khasawneh, 2005). The first study that empirically tested the relationship between training transfer and firm performance was published in 2014 by Saks and Burke-Smalley (2014). Results showed that training transfer mediates the path between training methods and firm performance.

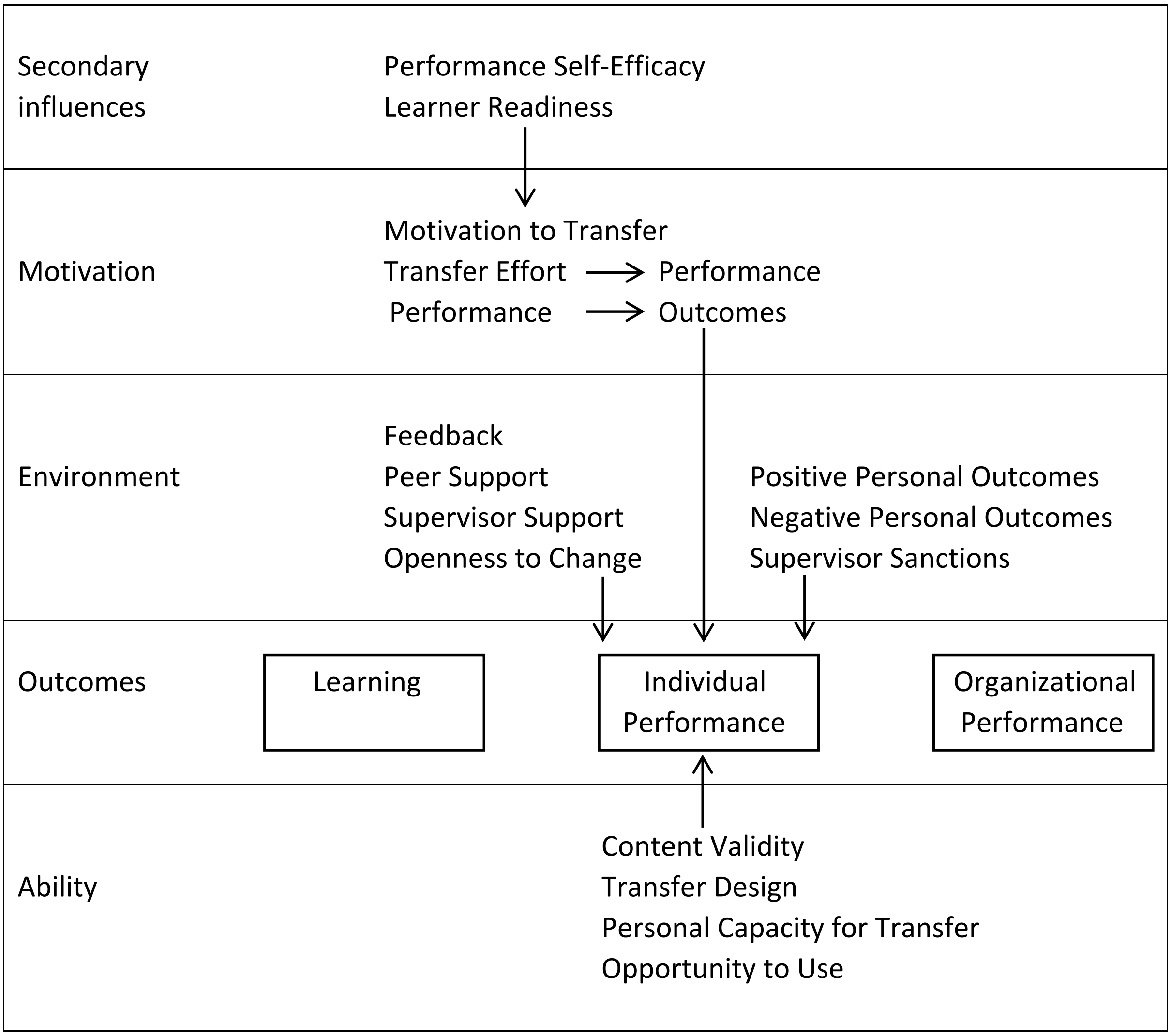

Figure 5. Learning Transfer System Inventory: Conceptual model of instrument constructs. Adapted from "Development of a generalized learning transfer system inventory" by E. F. Holton, R. A. Bates and W. E. A. Ruona, 2000, *Human Resource Development Quarterly, 11*(4), p. 339. Copyright 2000 by John Wiley and Sons. Adapted with permission.

Another theoretical framework that has received great attention in the training research literature is the transfer process model by Baldwin and Ford (1988) which is described next.

4.1.3 Transfer process model by Baldwin and Ford

Classical training evaluation research as well as companies primarily focus on measuring training success in the workplace in terms of practice-oriented outcome measures. In particular, two indicators are relevant to distinguish: on the one hand, *success of learning* that refers to immediate post-training knowledge and knowledge retention (Alliger et al., 1997); on the other hand, *success of transfer* that indicates lastin effects of the training and thus, plays a more crucial role in the long term (Park & Wentling, 2007). Training transfer is defined as “the degree to which trainees effectively apply the knowledge, skills, and attitudes gained in a training context to the job” (Baldwin & Ford, 1988, p. 63).

Often, companies observe that the knowledge and skills acquired in training are insufficiently transferred to the workplace (Grossman & Salas, 2011). For example, if the training situation differs greatly from the work situation, a phenomenon called *transfer problem* very likely occurs (Baldwin & Ford, 1988; Saks, Salas, & Lewis, 2014). This effect is confirmed by several scientific studies including meta-analyses (Grossman & Salas, 2011; Saks et al., 2014). For example, a survey by Saks (2002) indicated that 38% of trainees do not transfer immediately after training and this rises up to 56% after six months and up to 66% after one year. According to Saks (2002), this transfer problem is even underestimated in this study because these judgments of trainers might be inflated due to a self-serving bias. In addition, only about 50% of training investments result in an improvement in employees and the organization (Saks, 2002). In terms of training evaluation, it seems dangerous to evaluate training outcomes only immediately after training (Saks et al., 2014). Both success of learning and success of transfer are necessary to investigate. This raises the question: How are both constructs–success of learning and success of transfer–connected and how are they promoted? What do we know about influences on both constructs from a theoretical point of view?

Most research that systematically identified learning and actions for maximizing training effectiveness and their interactions is based on Baldwin and Ford’s (1988) transfer process model (see Figure 6). Reviewing 63 empirical studies, Baldwin and Ford (1988) investigated the relationship between *training inputs*, *training outputs*, and *conditions of transfer*. For a better understanding of factors influencing training outputs and transfer of learning (conditions of transfer), *training inputs* must be examined carefully. Training inputs predict training outputs directly and success of transfer indirectly through their impact on training outputs. In the transfer process model (Baldwin & Ford, 1988) training inputs are: (a) *trainee characteristics* (ability, personality, and motivation), (b) *training design* (principles of learning, sequencing, and training content), and (c) *work environment* (support and opportunity to use). Similarly to the distinction mentioned above, the model separately observes training outputs and conditions of transfer (Baldwin & Ford, 1988). Training outputs refer to the content learned (*learning*) and what is retained (*retention*) immediately after training. Transfer only occurs if the newly learned content is generalized and is consistently applied in the workplace. Thus, conditions of transfer refer not only to *generalization* but also to *maintenance* of targeted skills on a long term.

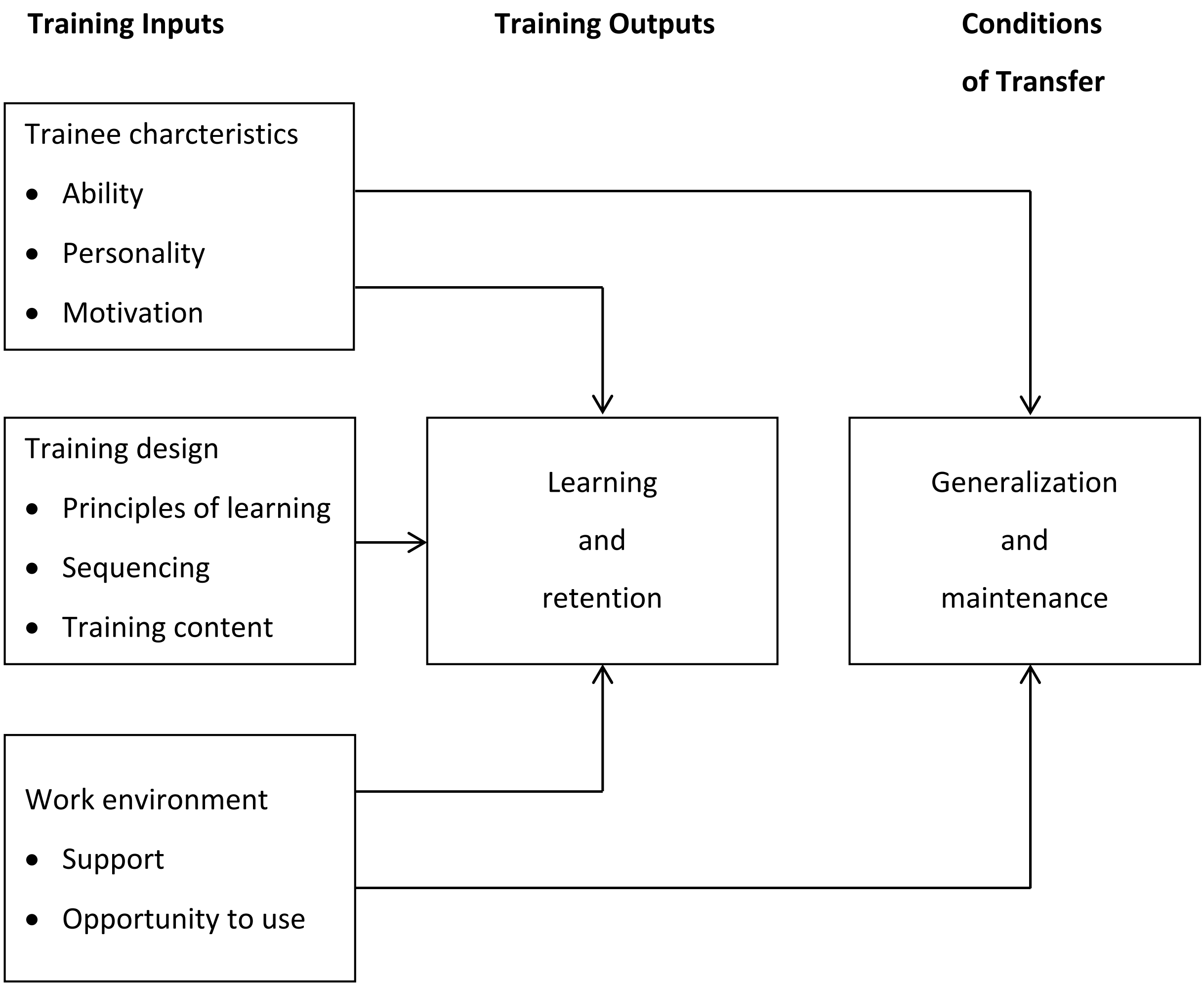

Figure 6. Transfer model by Baldwin and Ford (1988). Adapted from "Transfer of training: A review and directions for future research" by T. T. Baldwin and J. K. Ford, 1988, *Personnel Psychology*, *41*(1), p. 65. Copyright 1988 by John Wiley and Sons. Adapted with permission.

Since learning outcomes determine what content is potentially transferable to the workplace, success of learning is a necessary but not sufficient condition for success of transfer (Solga, 2011b). Without the existence of success of learning no success of transfer is possible. But according to Baldwin and Ford's (1988) model, as training inputs (trainee characteristics and work environment) also directly influence conditions of transfer, success of learning (training outputs) alone does not necessarily lead to success of transfer.

In recent decades, Baldwin and Ford's (1988) literature review received great attention indicated by over 2500 citations according to Google Scholar. Thus, their model served as a basis for further extensions (e.g., Baldwin, Ford, & Blume, 2009; Burke & Hutchins, 2007; Ford & Weissbein, 1997; Grossman & Salas, 2011). In sum, Baldwin and Ford's (1988) transfer process model has stimulated a vast amount of subsequent empirical studies that have identified when and for whom different training or transfer support methods are more or less effective and thus can best predict and improve training transfer (e.g., Burke, Hutchins, & Saks, 2013; Grossman & Salas, 2011; Salas et al., 2012; Saks et al., 2014).

4.2 Determinants of training success

Baldwin and Ford's (1988) model stimulated and thus succeeded in the training transfer literature (Saks et al., 2014). For example, the transfer process model was validated in a recent meta-analysis investigating 89 empirical studies. This meta-analysis confirmed that training transfer is influenced by a number of factors: namely trainee characteristics and work environment factors (Blume, Ford, Baldwin, & Huang, 2010).

As Baldwin and Ford's (1988) model is confirmed in numerous empirical studies (Saks et al., 2014), many other researchers adopted their categorization (e.g., Burke & Hutchins, 2007; Rouiller & Goldstein, 1993), sometimes using similar terms as in the distance learning literature: (a) trainee characteristics, (b) training design, and (c) interactions of the trainee with others (Lee, Driscoll, & Nelson, 2004). However, acknowledging the confirmation of Baldwin and Ford's (1988) well-established model but also taking theoretical models derived from the e-learning literature into account, the three major categories of training input determinants from Baldwin and Ford's model (1988) are applied in the following sections: (a) *trainee characteristics* (section 4.2.1), (b) *training design* (section 4.2.2), and (c) *work environment* (section 4.2.3). Here, specific determinants with their empirical contributions to maximize training success in different training settings, specifically in corporate e-learning when available, are discussed in closer detail.

4.2.1 Trainee characteristics

Trainees themselves are the central element within the learning process. Trainee characteristics (e.g., knowledge, skills, behaviors, and attitudes) serve as relevant determinants of training success. Therefore, trainee characteristics play an influential role in the learning and transfer process (Burke & Hutchins, 2007). Specifically *cognitive ability*, *training motivation*, *self-efficacy*, and *perceived utility of training* are identified as strong and most consistent factors enhancing training transfer (Grossmann & Salas, 2011). Moreover, research studies identified the *Big Five personality traits* (Barrick & Mount, 1991), *proactive personality* (Dura & Solga, 2010), and *demographic factors especially gender and age* as further influencing factors promoting training success (e.g., Colquitt et al., 2000). These factors and their empirical findings with regard to determine learning and transfer will be further chronologically specified.

Cognitive ability

A broad array of cognitive abilities is subsumed in the *g*-factor (e.g., Hunter, 1986; Jensen, 1986). As we know from a variety of research studies, *g* varies in individuals, for example, regarding the capacities of basic information processing or levels of cognitive resources (e.g., Kanfer & Ackerman, 1989). Consequently, when individuals differ in their learning process, they also differ in the acquisition of knowledge and skills. These differences can lead to different results in learning outcomes. Thus, cognitive ability predicts training success. The research literature has provided robust findings that intelligence or more precisely general cognitive ability strongly predicts training outcomes (e.g., Ree & Earles, 1991). Later, a broad meta-analysis comprising 20 years of training research summarized the antecedents of training outcomes and its results also indicated a strong relationship between cognitive ability and training transfer with a corrected correlation coefficient of r = .43 (Colquitt et al., 2000). More recently and also consistent with these previous meta-analytical findings, Blume et al. (2010) confirmed that cognitive ability is the strongest predictor for training transfer (for a literature review, see also Burke & Hutchins, 2007). Because individual differences in cognitive abilities are innate and therefore immutable, they can hardly be controlled by organizations. Nonetheless, those responsible for managing training must keep these individual differences in mind when assigning participants to a particular course of training. Moreover, it seems important to concentrate on variables that can be directly influenced to determine whether individual and situational characteristics explain any incremental variance in training outcomes.

Training motivation

In the traditional training research literature, motivation of trainees as a key variable in the learning process has been widely demonstrated (e.g., Baldwin et al., 2009; Lim & Johnson, 2002; Pugh & Bergin, 2006; Tziner, Fisher, Senior, & Weisberg, 2007). Training motivation can be defined as a process that includes intensity, direction, and persistence of learning-directed behavior (e.g., Kanfer, 1990) and it subsumes both *motivation to learn* as well as *motivation to transfer* (Gegenfurtner, Veermans, Festner, & Gruber, 2009). *Motivation to learn* is defined as the willingness to acquire new knowledge and skills within the course of training (Noe & Schmitt, 1986). But the framework of training evaluation rather focused on *motivation to transfer*, which is characterized by the trainee's effort to apply the new knowledge and skills to the workplace (Noe & Schmitt, 1986). The higher the motivation to transfer, the higher the likelihood that trainees apply the skills and knowledge learned (Baldwin & Ford, 1988). An empirical study investigating both constructs separately demonstrated a higher coherence between motivation to transfer and transfer (r = .43) compared to motivation to learn and transfer (r = .07); with interrelated motivation constructs of r = .26 (Chiaburu & Lindsay, 2008). The authors concluded that motivation to learn rather directly affects trainees' motivation to perform in the learning field, while motivation to transfer is more likely to trigger proactive behavior that is necessary for the actual transfer. However, the research literature has also demonstrated large differences in results regarding the relationship between motivation to transfer and success of transfer, ranging from r = .04 (Tziner, Haccoun, & Kadish, 1991) to r = .63 (Machin & Fogarty, 1997). Therefore, the mediating position of motivation to transfer is further discussed.

In the research literature, great importance has been attached to motivation to transfer as a mediator (e.g., Grohmann, Beller, & Kauffeld, 2014). Study results are presented in chronological order. Tannenbaum and Yukl (1992), for instance, found that the paths relating trainee characteristics or transfer climate to training success are mediated by motivation to transfer. These results were confirmed meta-analytically (Colquitt et al., 2000). In the model of Kontoghiorghes (2004), both motivation to learn and motivation to transfer, were detected as mediators. The mediating position of motivation to transfer was confirmed within an integrative literature review on past and current antecedents, correlates, and consequences of motivation to transfer (Gegenfurtner et al., 2009). The authors provided two exemplarily scenarios: On the one hand, no transfer of learning is given if the trainee is not motivated to apply the newly acquired knowledge and skills even though possible applications are provided. On the other hand, if no application is given on the job, transfer support actions seem practically impossible to implement, regardless of how high levels of motivation might be. In this case, the motivated trainee must actively look for other possible situations to apply the newly acquired knowledge and skills. A very recent study demonstrated that motivation to transfer predicts transfer (Grohmann et al., 2014). Also, motivation to transfer was identified as a mediational link between training characteristics (transfer design and perceived content validity) and transfer (Grohmann et al., 2014).

Prior to the training, motivation to transfer varies in trainees' individual characteristics (e.g., Big Five personality traits) and attitudes (e.g., concerns or targets). Colquitt et al. (2000) developed an integrative approach to training motivation. This approach highlights mediating variables (e.g., conscientiousness or self-efficacy) predicting motivation to transfer and success of learning. Pre-training signals or perceptions are also conducive to motivation to transfer (Baldwin & Magjuka, 1991). For example, trainees are more motivated to transfer what they have learned to the workplace when "they (a) received information prior to the training program, (b) recognized that they would have some accountability for learning with their supervisor, and (c) perceived a program as mandatory" compared to voluntary (Baldwin & Magjuka, 1991, p. 25). In contrast, a later meta-analysis found that voluntary participation is strongly related to transfer (Blume et al., 2010).

It is relevant to distinguish between the impact of *intrinsic* and *extrinsic motivation* on success of transfer (Burke & Hutchins, 2007). Trainees driven by intrinsic motivation to participate in training showed higher motivation in the course of training itself and also higher motivation to learn, both resulting in higher success of learning and success of transfer, while extrinsic motivation did not contribute significantly to success of learning and success of transfer (Facteau, Dobbins, Russell, Ladd, & Kudisch, 1995). In line with these findings, intrinsic motivation (r = .34) has a greater impact on remembering the content learned than extrinsic motivation (r = .05; Kontoghiorghes, 2001). However, transfer of learning is greatest when trainees were additionally motivated by extrinsic components that are anchored in their field of work (Taylor, Russ-Eft, & Chan, 2005).

Focusing specifically on e-learning settings, only a few studies shed light on the impact of motivation. Sitzmann, Brown, Ely, Kraiger, and Wisher (2009) proposed and empirically confirmed a cyclic motivation model for web-based training. Within a dynamic interplay, the model comprises different motivation constructs. The cycle starts with the participant's expectations on a specific e-learning course which have a positive impact on their motivation to learn. In turn, motivation to learn positively influences participants' reactions, which ultimately has a positive impact on their expectations for subsequent training. These results emphasize the relevance of a motivating training setting to develop or maintain a positive attitude towards e-learning courses. Summing up the findings above, the importance of trainee's motivation throughout the training and transfer process is empirically well-confirmed and widely replicated (see also Grossman & Salas, 2011; Grohmann et al., 2014).

Self-efficacy

Besides training motivation, self-efficacy as another trainee characteristic plays a decisive role in determining training success. Self-efficacy refers to the beliefs about one's own ability to perform specific behaviors. According to Bandura's (1986) social cognitive theory, self-efficacy is defined as the "belief in one's capabilities to organize and execute the courses of action required to manage prospective situations" (Bandura, 1995, p. 2). In the training context, self-efficacy refers to a person's belief in his or her ability to succeed in a particular training and later on in a transfer situation. Consequently, a trainee with a high level of self-efficacy has confidence in his or her abilities to achieve targeted objectives and will subsequently apply the content learned in the training (Colquitt et al., 2000; Saks, 1997).

The research literature meta-analytically has confirmed that pre-training self-efficacy strongly correlated with motivation to learn with $r = .42$ and transfer with $r = .47$ (Colquitt et al., 2000). Later studies consistently replicated positive correlations between self-efficacy and training transfer (e.g., Burke & Hutchins, 2007; Grossmann & Salas, 2011). For example, a mean correlation coefficient of $r = .22$ was meta-analytically demonstrated (Blume et al., 2010). Regression analyses showed that self-efficacy significantly predicts training transfer with $\beta = .30$ and $p < .001$ (Velada, Caetano, Michel, Lyons, & Kavanagh, 2007). In addition, self-efficacy is also positively correlated with training motivation (*r* = .34) which in turn significantly affects training transfer (Chiaburu & Lindsay, 2008; Chiaburu & Marinova, 2005). Finally, several meta-analyses demonstrated that self-efficacy and success of transfer are either directly or indirectly interrelated (e.g., Ford, Smith, Weissbein, Gully, & Salas, 1998; Holladay & Quiñones, 2003).

Although numerous studies showed a positive correlation, a meta-analysis showed that there are certain circumstances when high levels of self-efficacy seem counter-productive: Self-efficacy positively affects transfer performance only in tasks with low levels of difficulty but not in tasks with medium to high levels of difficulty (Judge, Jackson, Shaw, Scott, & Rich, 2007). Here, the expectancy of achieving the objective alone is insufficient; acquired knowledge and skills as well as relevant incentives play a rather decisive role (Bandura, 1977).

Focusing specifically on e-learning settings, external control is reduced compared to classroom training settings. Consequently, this individualized form of learning in particular requires self-regulated learning strategies to achieve learning objectives adequately and to coordinate the individual learning process. Trainees with a better knowledge and a more frequent use of self-regulated learning strategies were better in decoding and recoding knowledge and skills learned in a software training and they also showed higher levels of self-efficacy (Gravill & Compeau, 2008). Consequently, to promote learning outcomes in informal training settings which gain more and more importance in everyday work, self-tests assessing skills that need increased practice seem beneficial at various points in time within e-learning programs.

Perceived utility of training

The fourth presented trainee characteristic that promotes learning and transfer outputs is perceived utility. Trainees score higher in training transfer when they perceive a close connectedness between (a) their performance favored by supervisors and (b) their own valued and desired training outcomes (Chiaburu & Lindsay, 2008). Utility is more likely perceived when two conditions are met: (a) when trainees recognize a need to improve in their job performance and (b) when trainees actually believe that the content learned will certainly close their performance gap (Burke & Hutchins, 2007). Hence, knowing that trainees' perceived utility is crucial for training transfer, organizations benefit from emphasizing the necessity of training efforts and relevance to apply new skills to the workplace.

Big Five personality traits

In addition to the strong and most consistent factors enhancing training transfer identified by Grossmann and Salas (2011) as mentioned above, research studies investigated the Big Five personality traits and their relation to success of learning and success of transfer. Shortly after validating the five-factor model across instruments and observers by McCrae and Costa (1987) and adopting the model into personality research and assessment, a positive impact of selected Big Five personality traits was meta-analytically detected: Success of learning correlated positively with (a) conscientiousness, (b) extraversion, and (c) openness (Barrick & Mount, 1991). A later study confirmed that conscientious individuals are more likely to learn new knowledge and skills (Martocchio & Judge, 1997), however, a meta-analysis by Colquitt et al. (2000) found no evidence for this relationship. Instead, their analyses emphasized motivation to transfer and demonstrated a positive correlation between conscientiousness and motivation to transfer (r = .38). In 2002, another study confirmed that conscientiousness and success of learning did not correlate significantly (Herold et al., 2002). Instead, they identified openness to experience (r = .31) and emotional stability (r = .36) as essential dimensions for success of learning. However, a recent meta-analysis reviewing 89 empirical studies found that conscientiousness and success of transfer are moderately related (Blume et al., 2010). In summary, clear results regarding the impact of Big Five personality traits seem rather ambiguous and inconclusive.

Drawing attention to e-learning-specific training, a moderating role of the Big Five personality trait dimensions on the relationship between the trainees' level of control and performance was examined (Orvis, Brusso, Wasserman, & Fisher, 2010). Manipulating the extent of control (low or high), in learning environments with high levels of control (especially in video-based learning) trainees with high levels of openness and extraversion showed higher performance compared to trainees with lower levels of these personality traits. Conversely, trainees with low levels of openness and extraversion showed better performance in learning environments with low levels of control.

Proactive behavior

According to Dura & Solga (2010), proactive behavior must be considered when investigating training success. Proactive behavior is defined as an individual's initiative to organize and improve his or her own environment (e.g., Bateman & Crant, 1993; Crant, 2000). In other words, the term describes a proactive and self-determined approach to work ethic (Frese, Fay, Hilburger, Leng, & Tag, 1997). Due to a lack of direct applications of the knowledge and skills learned in the training or lack of necessary support from supervisors, trainees must actively seek for information and opportunities to transfer after training. Thus, they must show a strong proactive behavior to improve in the workplace (Dura & Solga, 2010).

Demographic factors: gender and age

As previously announced, especially gender and age are among the most investigated demographic variables when assessing training success (e.g., Colquitt et al., 2000). This meta-analysis showed that empirical studies on the effect of gender in the training context differ widely whereas the influence of age seems more stable. As anticipated, many studies demonstrated a negative relationship between age and success of learning (e.g., Colquitt et al., 2000; Gist, Rosen, & Schwoerer, 1988; Martocchio, 1994; Martocchio & Webster, 1992). In addition, it was found that age is negatively associated with the frequency of participation in training (Cleveland & Shore, 1992).

4.2.2 Training design

Referring to the transfer process model, training design is one of three major training input categories directly influencing training outputs and indirectly influencing transfer through their impact on training outputs (Baldwin & Ford, 1988; Lee et al., 2004). Meta-analyses demonstrated the effectiveness of the following training design concepts: behavior modeling (Taylor et al., 2005), error management (Keith & Frese, 2008), and realistic training environments (Kraiger, 2003). The training design and the manner of presenting information, particularly authenticity of the training environment and assigned tasks, play a key role in fostering success of learning and success of transfer (Grossman & Salas, 2011). From an instructional psychological point of view, the design of training settings is a key element for T&D professionals to actively and positively influence the learning process (Hochholdinger & Beinicke, 2011). Generally, when responsible T&D professionals set up a certain training, they have to choose precisely not only the content to be learned but also adequate training settings that consider the most effective principles of learning to promote success of learning in trainees (Velada et al., 2007). Back in the work environment, trainees will apply the newly acquired knowledge and skills more effectively if they have the feeling that the training was especially designed for the application to the actual work environment (Holton, 2005).

Within aspects of the training design, the trainer must carefully choose between e-learning, classroom training, or a combination of both. Research discussion on superiority of either e-learning or classroom training is presented in closer detail in section 4.3.1. Here, advantages and disadvantages of corporate e-learning are displayed by referring to meta-analyses and empirical studies (e.g., Means, Toyama, Murphy, & Baki, 2013; Sitzmann, Kraiger, Stewart, & Wisher, 2006).

However, with regard to the design of e-learning programs it is noteworthy to briefly present influencing constructs: namely *usability*, *age-adjusted e-learning designs*, and the *learner's control* with their empirical contributions.

Usability is defined as a measure of how easily and effectively trainees are able to handle the computer-based learning system to cope with learning tasks (Shackel, 2009). Instruments measuring usability are derived from research in software design and are based on standards, particularly the ISO-Norm EN ISO 9241. Perceived usability of a learning environment affects success of learning and consequently transfer of learning (Park & Wentling, 2007). In addition, the study showed that trainees completing an e-learning course with a positive attitude towards computers perceived higher levels of usability and thus assessed the training as more satisfactory and more effective. The authors concluded that these trainees also transfer the acquired knowledge in their daily work more effectively.

When designing training, it is essential to consider the impact of age and demographic change, especially to obtain or increase acceptance of corporate e-learning offerings. Thus, e-learning programs should be adapted to the learner’s age. A study by Bausch, Sonntag, Stegmaier, and Noefer (2010) illustrated that age-adjusted designs in e-learning programs lead to higher success of learning and success of transfer for young, middle-aged, and older trainees in both the short and long term. The study also demonstrated that the use of age-congruent role models in video-based learning plays a decisive role in training success.

Furthermore, when designing training, the importance of learner’s control has been demonstrated empirically by Fisher, Wasserman, and Orvis (2010). In their quasi-experimental study, trainees had the opportunity to complete an e-learning course with or without interactive control. Study results suggested that high levels of control promote affective reactions, and reactions that are beneficial to the learning process. In turn, these reactions have a positive impact on overall satisfaction with the e-learning program. Consequently, high levels of general satisfaction promote success of learning and success of transfer. Trainees with higher training satisfaction scores also provide more cognitive resources during training; consequently, they achieved higher levels of performance after training (Orvis, Fisher, & Wasserman, 2009).

4.2.3 Work environment

The third major training input category influencing training outputs directly and transfer both directly and indirectly through their impact on training outputs refers to the work environment (Baldwin & Ford, 1988). High levels of success of learning do not automatically generate high levels of success of transfer. According to the transfer problem, on-the-job learning must be considered in the context where knowledge and skills are applied. Eventually, as training interventions result in positive change towards targeted behaviors that increase work performance, trainees' ability to apply the newly acquired competencies to everyday work is the most crucial component of training effectiveness (Salas, Wilson, Priest, & Guthrie, 2006). Therefore, it is essential for responsible T&D professionals not only to meet the needs of trainees and implement adequate (e-learning) training programs but also to guarantee a work environment that supports the transfer process after training. Work environment subsumes (a) *transfer climate*, (b) *support from supervisors and colleagues*, (c) *opportunity to perform*, and (d) *follow-up actions*. According to Grossman and Salas (2011), these variables have shown the strongest, most consistent relationships with training transfer and therefore are presented here.

Transfer climate

Transfer climate is defined as all conditions in the work environment promoting or inhibiting training transfer (Rouiller & Goldstein, 1993; J. B. Tracey et al., 1995). In a positive transfer climate, for example, trainees' willingness to apply the content learned is higher compared to a negative transfer climate (Salas et al., 2006). A positive transfer climate is characterized as "cues that prompt trainees to use new skills, consequences for the correct use of skills and remediation for the incorrect or lack of use, and social support from supervisors and peers through the use of incentives and feedback" (Grossman & Salas, 2011, p. 112). According to Rouiller and Goldstein (1993), the combination of situational cues (e.g., manager goals, peer support, equipment availability, and opportunity to practice trained skills) and consequences (e.g., punishment and positive or negative feedback following the use of trained skills) determine transfer outcomes. Meta-analytical studies reported perhaps the strongest relationship (r = .37; Colquitt et al., 2000) between positive transfer climate and transfer among all other work environment factors (Blume et al., 2010). In contrast, a transfer climate with no support inhibited training transfer to the greatest extent as reported from the perspective of trainees (Gilpin-Jackson & Bushe, 2007), while a longitudinal study demonstrated that trainees' perceptions of sufficient practice, support from supervisors, the availability of documentation, and help from experts mediate the relationship between training and trainees' intentions to apply new skills to the workplace (Marler, Liang, & Dulebohn, 2006).

With special regard to a positive transfer climate in e-learning settings, positive correlations of facets of organizational learning culture (e.g., organizational framework of learning, aspects of T&D, or learning and developmental opportunities) with success of learning and success of transfer were found (Hochholdinger & Schaper, 2008). This particular study demonstrated that motivation to transfer mediates the relationship between organizational learning conditions with success of learning and success of transfer. Corporate structures as well as support from supervisors and colleagues are necessary for creating an ideal e-learning and transfer environment (Bates, Holton, Seyler, & Carvalho, 2000; Berge & Giles, 2008).

Support from supervisors and colleagues

At issue is the time frame and point of time when giving support to trainees. A temporary support from supervisors or colleagues immediately after completing training seems insufficient. All stakeholders including managers, trainers, and trainees must create and ensure that a supportive environment exists before and during training for efficient and effective learning, and to later maximize transfer of learning.

Another issue of supporting trainees considers the person or group of individuals providing support such as supervisors or colleagues. A meta-analysis demonstrated that the relationship between support from supervisors and transfer of learning is stronger with $r = .31$ than the correlation between support from colleagues and transfer of learning with $r = .14$ (Blume et al., 2010). The authors noted that these results are based on small sample sizes and thus, results must be interpreted with caution. However, in terms of providing feedback, no superiority of either feedback from supervisors or feedback from peers was found. Instead, the number of feedback providers as well as the helpfulness of the feedback were positively connected to motivation to transfer and consequently to the actual training transfer (Van den Bossche, Segers, & Jansen, 2010). When considering social support (supervisor or coworker) or organizational support as moderator in the relationship between perceived learning and perceived transfer, this relationship was stronger when trainees had coworker support (Homklin, Takahashi, & Techakanont, 2014). Generally, the research literature has focused on supportive behaviors from (a) supervisors and (b) colleagues separately and therefore, this scheme is adopted in the following.

First, supportive behaviors and actions from supervisors have multiple facets and can facilitate learning processes before, during, and after training. It has been demonstrated that the following supportive behaviors and attitudes from supervisors are the most supportive determinants of training transfer: information sharing, direct feedback, and the provision of resources (Awoniyi, Griego, & Morgan, 2002; Kontoghiorghes, 2001). Also, it seems effective and easily manageable by supervisors if they demonstrate recognizing, rewarding, or modeling behaviors (Salas et al., 2006). With an integrative literature review on training transfer, Burke and Hutchins (2007) revealed transfer support actions from supervisors including discussions with the trainee on the acquired knowledge in the training, participating in training themselves, providing support, and coaching trainees. In the eyes of trainees, speaking of acquired knowledge with the supervisor, involvement of the supervisor in the training, and positive feedback from the supervisor were perceived as most important supportive factors (Lim & Johnson, 2002). Participation of supervisors themselves was strongly associated with post-training utilization (Gilpin-Jackson & Bushe, 2007).

Regarding possible consequences or effects of supportive behaviors from supervisors, Cromwell and Kolb (2004) demonstrated that trainees with high levels of active supervisor support transferred more knowledge and skills one year after training compared to those with low levels of supervisor support. In addition, precise short- to long-term goal settings sought by both supervisor and trainees can promote transfer of learning by drawing attention to the desired goal, which can stimulate and increase trainees' application of the newly acquired knowledge and skills (Burke & Hutchins, 2007; Locke & Latham, 2002). In combination with feedback, specific and difficult goals trigger motivation and consequently improve performance (Robbins & Judge, 2009).

Second, supportive behaviors and actions from colleagues (e.g., reciprocal observation, add-on knowledge exchange, or communicating and sharing ideas about training content) directly affect trainees' transfer of learning (Gilpin-Jackson & Bushe, 2007; Hawley & Barnard, 2005). Similarly to positive effects of supervisor support, active peer support also resulted in a positive transfer effect even one year after the actual training took place (Cromwell & Kolb, 2004). Analyses with structural equation modeling demonstrated that contextual factors (e.g., peer support) predicted pre-training motivation which in turn was related to skill transfer (Chiaburu & Marinova, 2005). This positive relationship between support from colleagues and transfer of learning was also found meta-analytically (Blume et al., 2010). Contrary to expectations, the recent study by Homklin et al. (2014) mentioned above demonstrated that neither organizational nor supervisor support but only coworker support moderated the relationship between learning and transfer.

Opportunity to perform and follow-up actions

Trainees need resources and opportunities to apply their newly acquired knowledge and skills to the workplace (e.g., Burke & Hutchins, 2007). Referring to the transfer process model, aside from social support, opportunity to use is highly valued by trainees (Lim & Johnson, 2002) and greatly influences training transfer (Baldwin & Ford, 1988). But in what way can organizations or supervisors provide opportunities? First, to achieve positive transfer outcomes, opportunity to use should follow closely after the training course (Salas et al., 2006). Second, organizations can provide opportunities in terms of setting times and supplying resources for the utilization of new skills (Cromwell & Kolb, 2004; Gilpin-Jackson & Bushe, 2007). Third, supervisors can coordinate current workloads similar to current training contents, so that trainees can immediately practice and transfer what they have learned (Clarke, 2002). If such opportunities are not provided by the organization, the transfer process is likely to fail.

The research literature has emphasized the important role of so-called follow-up actions referring to post-training interventions (Baldwin et al., 2009). Such interventions include reflections on training experiences, repeated practice sessions, follow-up training, add-on discussions, job-aid tools, and feedback loops that continuously facilitate the learning process (Hattie & Timperley, 2007). Overall, a supportive work environment is crucial for the application and maintenance of new skills and should be considered by T&D professionals.

4.3 Evaluating training settings in practice: shift from classroom training to e-learning

Since training evaluation in various career development processes in later career stages is a widely diverse research topic, it seems relevant to concentrate on current trends or discussions in research and practice. Two issues are suggested: identifying and evaluating the trend.

The current trend in training evaluation research refers to the shift from traditional classroom training to e-learning in corporate practice. In view of the fact that training settings applied in practice vary greatly across educational settings and institutions, two training settings are fundamentally distinguished: e-learning and classroom training. Generally, transfer of information via e-learning settings has become more important in recent decades. As e-learning programs are widely implemented in the corporate sector, companies experience a shift from traditional classroom training to e-learning. Statistics describing this shift demonstrate that the companies that continue to innovate in corporate e-learning settings are considered as market leaders. For example, according to a U.S. study in 2001, 80% of the U.S. Fortune 500 companies already implemented e-learning programs or planned to do so in the near future (Hammond, 2001). Similarly in Germany, the majority of Germany's major enterprises implemented e-learning programs for further training in addition to traditional classroom training. In 2010, the MMB-Institute for Media and Competence Research (MMB) conducted a study that demonstrated that 55% of the top 500 companies in Germany implemented e-learning settings; with web-based training, computer-based training, and blended learning being among the most implemented forms of corporate e-learning (MMB, 2010). Up to now, such e-learning settings have become a standard procedure in the corporate training context (Hochholdinger & Beinicke, 2012; MMB, 2010; Rey, 2009).

Evaluating this current trend, advantages and disadvantages of corporate e-learning in the research literature are presented (section 4.3.1). Then, from the state of research literature, the next section 4.3.2 gives answers to the question of how to be effective in corporate e-learning and classroom training as a conclusion.

4.3.1 Advantages and disadvantages of corporate e-learning

Before revising advantages and disadvantages of corporate e-learning, clarifying terminology seems relevant: The terms web-based training, e-learning, or *corporate e-learning* (connected with companies) are often used synonymously (Newton & Doonga, 2007; Rey, 2009). *Corporate e-learning* refers to corporate training and educational programs that subsume a variety of learning activities that allow trainees to gain knowledge and skills using information and communication technologies (e.g., Han, Dick, Case, & Van Slyke, 2009). Often, various learning activities are delivered to trainees through a variety of technical tools such as learning management systems, course management systems, intranets, the internet, or stand-alone modules (Dittler, 2011). Such learning platforms do not only provide contents of e-learning courses, but they are also able to manage data of trainees such as performance results, certificates, and access privileges (Dittler, 2011).

Advantages

The rapid obsolescence of knowledge and the huge demand for cost-efficient and effective training programs are primary reasons for the increasing implementation of e-learning settings in companies (Fry, 2001). From the perspective of T&D professionals, an exploratory study revealed that e-learning is most valuable for delivering instructions governing familiar company tasks such as providing information about products, fulfilling compliance requirements, and securing standardizations (Rossett & Marshall, 2010).

In addition, a rather detailed overview of multiple motives for implementing web-based training from the perspective of training providers is summarized by Newton and Doonga (2007) (see Table 19). While external training providers emphasize the quality of learning materials and the learner's motivation, both of which expected to lead to better work performance and higher satisfaction, internal T&D professionals highlight the flexibility of training times and reduced times away from work.

Table 19. *Overview of Multiple Motives for Implementing E-Learning Programs from the Perspective of Training Providers*

1	Adequacy of the technology to support remote learning
2	Ability to deliver training anywhere, anytime, and to anyone
3	Cost savings due to elimination of expensive travel costs
4	Just in time training because of access to timely information
5	Higher retention of content by learners because of personalized learning
6	Able to deliver compliance training much more effectively
7	Ability to monitor learner progress to ensure completion
8	Improved collaboration and interactivity amongst students
9	E-learning is more “risk free” and less intimidating than instructor led training
10	Trainees prefer to use e-learning
11	Ability to incorporate simulations/games/stories to make the learning more interesting

Note. Adapted from Newton and Doonga (2007).

Similarly to Newton and Doonga’s (2007) overview of motives, Macpherson, Elliot, Harris, and Homan (2004) reviewed the core advantages and disadvantages of corporate e-learning in practice listed in the research literature. The most persuasive advantages are flexibility and cost savings. In the following, the three facets of flexibility and their consequences are further reviewed. Then, advantages of cost savings are summarized briefly.

Corporate e-learning is flexible concerning (a) location, (b) time, and (c) adjustment to individual needs, prior knowledge, and individual interests (Goldstein & Ford, 2002; Welsh, Wanberg, Brown, & Simmering, 2003). First, corporate e-learning is not bound to a specific location; consequently, travel expenses are reduced or eliminated. In combination with the World Wide Web, learning contents can easily be made available for a larger number of employees at once, independent of their current location, which is crucial, especially for international and global companies (Hochholdinger & Beinicke, 2012; Macpherson et al., 2004). Second, corporate e-learning is not time-bound; thus, access to information at any time allows more flexibility in the workplace. Corporate e-learning gives the possibility of providing globally consistent training that is available just-in-time. Third, corporate e-learning is flexible in the pace and in the learning curve of each individual. Especially from an educational point of view, it is important and beneficial that the learning pace can be adapted according to individual capability and circumstances (Caudron, 1999; Sandelands & Wills, 1996). Thus, corporate e-learning can not only provide adapted content, but it can also give feedback that evidently can increase the motivation of trainees (Pollard & Hillage, 2001;

Schriver & Giles, 1999). All three facets of flexibility aim at resulting in higher levels of trainee satisfaction and cost savings (e.g., travel costs).

Advantages are often referred to cost savings not only in terms of reduced training times as training can be provided for a large number of trainees at once but also in terms of reduced travel expenses and reduced times away from the job (Schriver & Giles, 1999). Living in a world of economic competition, ROI is considered as a good indicator of economic success. This concept is applied to the training context, too. S. Swanson (2001) investigated a study demonstrating that 46% U.S. businesses already see a return on their investments in training, while 94% expect to see returns or further returns within the next two years. Further, 80% of the U.S. Fortune 500 companies use or intend to use e-learning settings expecting a significant ROI (Hammond, 2001). The company IBM serves as a good example of a large U.S. Fortune 500 company with a strong learning culture that demonstrates significant of cost savings: In 1999, IBM reported savings of $200 million when they implemented e-learning programs as their primary training delivery setting (Tai, 2005). Another example is Ernst and Young that reported savings in reduced training expenses of 35% when they moved 80% of their training online (Newton & Doonga, 2007). However, U.S. experts predicted that the U.S. e-learning market would hit $16.7 billion in 2009, and will grow to $23.8 billion by 2014 ("US eLearning Market Reaches $16.7 Billion in 2009", 2009, para. 1). In summary, the most persuasive advantages of corporate e-learning refer to flexibility and cost savings.

Disadvantages

Macpherson et al. (2004) argue that disadvantages are largely ignored, not only in practice but also in the research literature. An overview of possible difficulties in response to supposed strengths when implementing e-learning programs is briefly summarized in Table 20 (Hochholdinger & Beinicke, 2012).

Table 20. *Overview of Possible Difficulties in Response to Supposed Strengths when Implementing E-Learning Programs*

Supposed strengths	Possible difficulties
Low operating costs, cost efficiency	High investment and operating costs
Fast access, easy distribution	High standards/needs of technical infrastructure
Flexibility in location and time	Difficult compatibility of training and working hours
Motivating, preferred training setting	Low acceptance, high drop-out rates
Self-regulated, autonomous learning	Intensive support/supervision

Note. Adapted from Hochholdinger and Beinicke (2012).

On the one hand, the implementation of e-learning programs is associated with high investment costs. However, on the other hand, the implementation can also cause continuing high operating costs for administration and information technology. Technology-related difficulties can occur if, for example, smooth access to the internet is not guaranteed or trainees lack in the technical know-how to initiate the learning process (Sitzman et al., 2006). From the perspectives of trainees, skills for technology management and self-directed learning are necessary for effective learning (Dringus, 2000). In some cases, it is difficult to incorporate training with working hours. A negative trend regarding the acceptance towards e-learning has also been described in the research literature (Bürg & Mandl, 2005). If acceptance towards e-learning is lacking or the company's learning culture is insufficient, self-regulated corporate e-learning can cause high drop-out rates (Hochholdinger & Beinicke, 2012). In addition, e-learning provides limited opportunities for individual feedback (Dittler, 2011). In summary, despite high investment and operating costs, corporate e-learning can lack in high standards of technical infrastructure, compatibility of training and working hours, support for trainees, or adequate evaluation processes.

4.3.2 Training effectiveness in corporate e-learning and classroom training

The current literature discussion on advantages and disadvantages of corporate e-learning raises the question: Which training setting–e-learning or classroom training–is most appropriate given certain circumstances? In recent decades, several studies addressed this question (e.g., Sitzmann et al., 2006). Therefore in the following, primarily meta-analytical findings on the effectiveness of e-learning (also referred to web-based instruction or distance education) and classroom training (also referred to classroom instruction or face-to-face education) are presented at a summary level.

Though this issue has been questioned frequently, the overall effectiveness and thus the importance of corporate training has been confirmed (Arthur, Bennett, Edens, & Bell, 2003). Results of this meta-analysis revealed medium to large effect sizes for corporate training regarding Kirkpatrick's (1959) evaluation features: mean effect sizes of d = 0.60 for reaction criteria, d = 0.63 for learning criteria, d = 0.62 for behavioral criteria, and d = 0.62 for results criteria (Arthur et al., 2003). However, results focusing on the relationship between various evaluation criteria and specific training settings (e-learning or classroom training) varied widely.

No differences between e-learning and classroom training. A meta-analysis on the effectiveness of distance education by Zhao, Lei, Yan, Lai, and Tan (2005) reported no significant difference in outcomes between distance and face-to-face instructions (for previous meta-analyses, see also Cavanaugh, 2001; Machtmes & Asher, 2000). Accordingly, a later meta-analysis by Sitzmann et al. (2006) found no significant differences in outcomes when instructional methods (e.g., a presentation) were applied in both training settings and no differences were found in trainee satisfaction.

Classroom training dominating e-learning. A meta-analysis showed that few studies investigating both training settings revealed smaller effect sizes for e-learning within the categories reactions (d = 0.31), behavioral changes (d = 0.32), and success of learning (d = 0.40) compared to classroom training (Arthur et al., 2003). Trainees reacted somewhat more favorably towards classroom instruction than blended instruction (Sitzmann et al., 2006).

E-learning dominating classroom training. However, the authors demonstrated that results differ depending on the type of content learned: Web-based and classroom instructed training did not differ significantly when teaching procedural knowledge (Sitzmann et al., 2006). However, web-based instruction was somewhat more effective than classroom instruction when teaching declarative (factual) knowledge, especially in long courses (Sitzmann et al., 2006). This superiority effect of web-based training increased by 19% when (a) trainees were able to control the content and sequence in the learning process, (b) trainees received feedback during training, and (c) trainees had opportunities to practice the training material. Similarly, a more recent meta-analysis and literature review of online training studies from 1996 to 2008 revealed that students performed modestly better in e-learning settings than students learning the same content through face-to-face instructions (Means, Toyama, Murphy, Baki, & Jones, 2010). However, the authors claimed that training programs varied to a great extent in the valence of outcomes in both training settings.

This raises the question: Under which circumstances is corporate e-learning most effective? To be effective, corporate e-learning must be built upon established learning theories that are applicable to the adult learning environment. In coherence with traditional classroom training, corporate e-learning must be based on specific assumptions and principles to have an effect (Bedwell & Salas, 2010). Within the e-learning program, it is important that the content learned is (a) demonstrated vividly, (b) practiced deliberately and frequently by trainees, and (c) supported by feedback. For example, observational learning, as one technique suitable to demonstrate complex sequences of behaviors, can be realized by watching videos and animations (Hochholdinger & Beinicke, 2012). Deliberate practice requires repeated queries in the form of self-tests including evaluations. Such evaluation outputs identify content already mastered and content which needs further practice. In addition, trainees must independently regulate their learning process to adapt their learning pace and individual level of difficulty (Sitzmann et al., 2006). When implementing *new* learning technologies, trainees must be prepared to learn with the *new* way of interaction. Therefore, it is particularly important that the training software and specifically the learning program is pleasant and efficient for trainees to use (Hochholdinger & Beinicke, 2011).

Frequently underestimated, it is crucial to update and adapt learning platforms with their contents and technical tools applied (Hochholdinger & Beinicke, 2012). Considering such beneficial learning principles, research studies showed that e-learning is effective and can dominate traditional classroom training (e.g., Sitzmann et al., 2006). Strengths and weaknesses in both e-learning and classroom training can lead to success or failure.

Still, the question remains whether positive effects in learning outcomes are elicited when elements of online and face-to-face conditions are combined or blended. Learning outcomes were larger when online and face-to-face elements were blended compared to purely online or purely face-to-face instruction (Means et al., 2010). Similarly, Sitzmann et al. (2006) meta-analytically found that a blended instruction was 13% more effective than classroom instruction for teaching declarative knowledge (d = 0.34) and 20% more effective than classroom instruction for teaching procedural knowledge (d = 0.52). Research studies in the field of higher education point out that using blended learning promotes independent and community learning at the same time (Garrison & Kanuka, 2004). It is the face-to-face element in blended learning that sustains trainee commitment and reduces possible isolation of purely web-based learning in higher education (Wall, Ahmed, & Smit, 2006). Furthermore, blended learning can promote learning communities, provide follow-up resources in communities of practice, provide access to guest experts, and provide mentoring or coaching via face-to-face or online activities in higher education and workplace settings (Bonk, Kim, & Zeng, 2005). However, the advantage for blended learning is not necessarily based on the type of media per se, but it is rather due to differences in content, pedagogy, and learning time that were more conducive in blended conditions (Means et al., 2010). Therefore, it seems important for blended learning settings to adequately match the choice of the training delivery setting with the didactical design of a blended learning arrangement (content, communication, construction) to ensure that the actual learning objective is met (Kerres & De Witt, 2003; Salas et al., 2012).

In summary, opportunities and threats must be weighted up against each other to achieve the main objective of training: the sustainable and effective transfer of the learned content to the workplace (Sitzmann et al., 2006). It is particularly important to plan training activities effectively and to evaluate their benefits (Salas et al., 2012). A solid training evaluation helps to monitor learning and transfer processes, especially when new e-learning programs are implemented. More research is needed regarding the effectiveness of different learning conditions. It is equally essential to investigate experimental studies including control groups and measuring learning and transfer outcomes. Unfortunately so far, there is only few empirical research on systematic evaluations regarding the success of learning and success of transfer of specific corporate e-learning programs. For the lack of empirical research on determinants promoting success of learning and success of transfer in corporate e-learning, Shmikler (1996) notes the following key reasons: non-specific learning objectives, vague purpose of the training, unrelated training factors influencing trainees' performance, lack of management support, insufficient opportunities for training, and inadequate rewards for transferring new skills.

To bridge this gap, within the framework of training evaluation, empirical evidence is presented by examining specific outcome measures such as success of learning and success of transfer as well as their determinants. Subjective training success measures from trainees themselves as well as objective training success measures (performance test) from trainees were assessed.

CHAPTER 5 EMPIRICAL EVIDENCE FROM STUDIES ON TRAINING EVALUATION

General research trend: shift from classroom training to e-learning in corporate practice

Training is often part of career development processes in later career stages. Research in the field of training evaluation often aims at measuring training success especially focusing on success of learning and success of transfer as well as their influencing constructs. So far, research in the field of training evaluation primarily focused on classroom training as the training setting of choice. However, the degree of virtuality in everyday life and work life has increased considerably in recent years. Thus, in the field of training, virtual training settings have been increasingly implemented in companies. In fact, they have partially complemented or even replaced traditional classroom training (Hochholdinger & Beinicke, 2012; MMB, 2010). In the course of new media services, we face the question of how effective e-learning is.

In this dissertation, three empirical studies on training evaluation combine the following issues: measuring training success and its determinants in e-learning and/or classroom training in corporate settings. Training programs are often evaluated by focusing on adult employees in the work environment. Compared to other studies, a new perspective is chosen here: Both employer and employees are asked to evaluate training success. Thus, this dissertation aims at integrating both the organizational level (meso-level) and the individual level (micro-level). To be more precise, this research explores (a) the employers' view, especially T&D professionals' view, of evidence-based actions for maximizing training effectiveness (Study 4) as well as (b) the employees' view of their own training success (Studies 5 and 6). Each study focuses on specific research gaps with their associated research questions, which are briefly summarized in the following.

First specific research gap: application of evidence-based actions for maximizing training effectiveness in corporate practice (Study 4)

To optimally use the training contents learned, research in the field of T&D has identified a variety of actions for maximizing training effectiveness (Salas et al., 2012). But to what extent do corporate businesses apply evidence-based actions to maximize training effectiveness (all of them well-investigated in the current research literature of T&D)? How intensive do companies already support their e-learning and classroom training by implementing actions for maximizing training effectiveness? Concerning this matter, are there differences between e-learning and classroom training? Are there differences in the implementation whether actions for maximizing training effectiveness are applied before, during, or after training? These questions are addressed in the fourth study, as this study investigated the extent to what evidence-based actions for maximizing training effectiveness are already implemented before, during, and after e-learning and classroom training in companies that have been awarded as "best employers in Germany" (section 5.1).

Second specific research gap: comparison of training success between corporate e-learning and classroom training (Study 5)

Training success is based on the application of training content to the workplace. In addition to classroom training, further education in the form of e-learning is increasingly used in the corporate context due to technological progress. However, this development is not considered sufficiently in training evaluation research. In particular, Aguinis and Kraiger (2009) claimed that effectiveness studies examining predictors of success of learning and success of transfer as facets of training success are needed, thus influencing training effectiveness specifically in e-learning settings. So far, however, there are only few empirical studies that systematically compare training success in both training settings. Are there differences between e-learning and classroom training regarding objective as well as subjective training success? Do training setting-specific patterns of training success change over time? As the fifth study provides answers to these questions, it contributes to this specific research gap. The fifth study explored differences in training success of trainees in a specific corporate e-learning and classroom training using a field experiment with a time-lag design (section 5.2).

Third specific research gap: determinants of training success specifically in corporate e-learning settings (Study 6)

Based on Baldwin and Ford (1988), current approaches of training evaluation research increasingly concentrate on global models that include both the individual and the organizational framework of training success (Kauffeld, Brennecke, & Strack, 2009).

Therefore, many studies examine not only the success of learning and success of transfer of trainees per se but also the role of interacting variables derived from both the individual and the organizational level. On the individual level, variables describing trainee characteristics (e.g., motivation to transfer, personal capacity for transfer, self-efficacy) and on the organizational level, variables referring to training design and work environment are focused (e.g., transfer design, supervisor support, performance coaching). In recent decades, T&D practice has increasingly tended to shift from traditional classroom training to e-learning. So far, the research literature has shown only few empirical studies on determinants of training success specifically in corporate e-learning settings (Aguinis & Kraiger, 2009). What are key factors in predicting or promoting success of learning and success of transfer in corporate e-learning settings? The sixth study provides answers to this question as this study empirically examined determinants of promoting training success in corporate e-learning settings (section 5.3).

Chapter outline

Contributing to the literature of career development in later career stages, current research questions or issues in the field of training evaluation as mentioned above are addressed in chapter 5. Each of the three studies on training evaluation is separately presented according to APA standard research format: aims and hypotheses of the study; methodology including a description of participants, instruments, and procedures; results; and discussion. Effect sizes were calculated using the psychometrica website (Lenhard & Lenhard, 2014). Within the discussion, study results are summarized and embedded within the research literature. Limitations of each study and implications for future research and practice are discussed.

5.1 Evidence-based actions for maximizing training effectiveness in corporate e-learning and classroom training

5.1.1 Aims and hypotheses

Aims

The study aimed at investigating the transfer problem in the workplace. Thus, the study focused on the perspective of T&D professionals employed at companies which are diverse in branches of industry and among the best and most popular employers in Germany. The selection of the companies was based on rankings of best workplaces in Germany awarded by three research and consulting institutes.

On the whole, the study investigated various issues: First, general information about implementing e-learning, such as period and various forms, was assessed. Second, the study identified which topics were addressed and which trainees participated in corporate e-learning and classroom training. Third, as to the main research focus, the study examined the implementation of actions for maximizing training effectiveness for e-learning and classroom training settings. Thus, the study investigated the extent to what actions for maximizing training effectiveness that are well-known and well-evidenced were implemented in both training settings before, during, and after training. Finally, the question arose whether there were differences in the implementation of actions for maximizing training effectiveness between e-learning and classroom training across three time periods.

Concerning practical contributions, on the one hand, study results were provided as a benchmark, and on the other hand, a checklist of evidence-based recommendations and best practices for maximizing training effectiveness was provided to the companies that are among the best employers in Germany, thus fostering the success of transfer in the workplace.

Hypotheses

Implementation of e-learning and classroom training. In view of the fact that training settings applied in practice vary greatly across educational settings and institutions, the way the learning content is delivered to trainees can be broadly divided in two training settings: e-learning and classroom training. Generally, when responsible T&D professionals set up a certain training, they have to choose the adequate training setting in precise accordance to the content to be learned and to consider the most important principles of learning in order to promote trainees' success of learning (Velada et al., 2007). Back in the work environment, trainees will apply newly acquired knowledge and skills better if they get the feeling that the training was especially designed for the application in the actual work environment (Holton, 2005). Within aspects of the training design according to Baldwin and Ford's model (1988), the trainer has to carefully choose either e-learning, classroom training, or a combination of both.

A criterion of selecting the adequate training setting is the topic or type of content learned. E-learning settings are predominantly applied in the area of declarative (factual) knowledge such as hard skills, whereas classroom instructed training was rather implemented when teaching procedural knowledge such as soft skills (MMB, 2010). Further regarding the effectiveness, a study by Sitzmann et al. (2006) revealed that web-based instruction was somewhat more effective and more frequent implemented than classroom instruction when teaching declarative (factual) knowledge. Therefore, we postulate:

Hypothesis 1: *Hard skills are more frequently trained in e-learning settings, whereas soft skills are more frequently trained in classroom training.*

Implementation of e-learning-specific actions for maximizing training effectiveness. In the last decades, transfer of information via e-learning has become more and more important, which is indicated by a vast amount of implementations of e-learning settings in the corporate sector (Fry, 2001). Moreover, as previously presented in section 4.3, companies even experience a shift from traditional classroom training to e-learning. According to the transfer process model (Baldwin & Ford, 1988), training design plays a key role in training transfer. Thus, regardless of which training setting is used (e-learning or classroom training), it is important to construct training designs that highly support training transfer. Specifically for e-learning settings, Salas et al. (2012) identified e-learning-specific actions for maximizing training effectiveness. Responding to the current shift from traditional classroom training to e-learning, the question arises to what extent do corporate businesses apply these e-learning-specific actions to maximize training effectiveness. As T&D professionals are responsible for guiding the learning process of employees actively and maximizing success of learning and success of transfer (Hochholdinger & Beinicke, 2011; Velada et al., 2007), their perspective was investigated.

Research question: *To what extent do companies implement e-learning-specific actions for maximizing training effectiveness that maximize the success of learning and success of transfer?*

Implementation of actions for maximizing training effectiveness comparing training settings. The next issue refers to the comparison between training settings regarding the implementation of actions for maximizing training effectiveness separately for each time period (before, during, and after training). To optimally use the training contents learned, research in the field of T&D has identified a variety of actions for maximizing training effectiveness (Salas et al., 2012). But to what extent do corporate businesses apply evidence-based actions to maximize training effectiveness (all of them well-investigated in the current research literature of T&D)? How intensive do companies already support their e-learning and classroom training by implementing actions for maximizing training effectiveness? Furthermore, are there differences between e-learning and classroom training?

Companies invest in e-learning settings to save costs (Bedwell & Salas, 2010). Consequently, professional consulting for e-learning programs is barely sought due to costly investments of time and money (i.e., regarding drafts of e-learning tools). Less than 1% of the costs invested in e-learning settings were spent on consulting (i.e., for implementation, didactical design, or application in practice) for professional e-learning programs supporting transfer of learning (MMB, 2010). Moreover, it was claimed that companies are rather interested in offering innovative and entertaining learning platforms to their employees instead of focusing on its potential regarding transfer of learning (Bedwell & Salas, 2010). Therefore, we propose:

Hypothesis 2: *Less actions for maximizing training effectiveness are implemented in e-learning compared to classroom training settings across all three time periods–before (Hypothesis 2a), during (Hypothesis 2b), and after training (Hypothesis 2c).*

Implementation of actions for maximizing training effectiveness comparing time periods. The last issue refers to the transfer problem itself, which has been demonstrated in several scientific studies including meta-analyses (e.g., Baldwin & Ford, 1988; Grossman & Salas, 2011; Saks et al., 2014). That is, fewer actions for maximizing training effectiveness are implemented after training courses took place. For example, support from supervisors and colleagues, positive and negative consequences, phases of exercises, meaningful and analytical feedback are transfer support actions after training when applying the context learned, but they are often insufficiently considered. In the corporate practice, only 9% of the companies assessed success of transfer and only 7% assessed organizational results within their evaluations (Van Buren & Erskine, 2002). Instead, companies investigated success of satisfaction with 78% and success of learning with 32% in their evaluation (Van Buren & Erskine, 2002). Further, companies tend to focus on aspects during training by presenting factual information and nice-looking designs (Salas et al., 2012). Thus, we hypothesize:

Hypothesis 3: *Actions for maximizing training effectiveness are least implemented after training compared to before and during training in both e-learning and classroom training.*

5.1.2 Methodology

Participants and Procedure

To gain a representive insight the work environment, our sample comprised T&D professionals of companies which are among the best and most popular employers in Germany. The selection of the sample was based on rankings of best workplaces in Germany awarded by the *Top Employers Institute*, *Great Place to Work Germany*, and *Trendence Institute*–three research and consulting institutes (see Table 21).

Table 21. *Research and Consulting Institutes Awarding Best Workplaces in Germany*

Institute	Category
Top Employers Institute	Top Employers Germany
	Top Employers Engineers Germany
	Top Employers Automotive Germany
Great Place to Work Germany	Germany’s Best Workplaces 2012
	Best Workplaces in Health Services
Trendence Institute	Business
	Engineering
	Information Technology
	Law

The ranking of the *Top Employers Institute* comprised the categories Top Employers Germany, Top Employers Engineers Germany, and Top Employers Automotive Germany and included about 170 companies (Top Employers Institute, 2013). Evaluation criteria of the Top Employers Institute were especially T&D, career opportunities, and the companies’ culture. *Great Place to Work®* is a global human resource (HR) consulting, research, and training company that is specialized in large-scale assessment (Great Place to Work, 2014a, para. 1). Its annual global research data represent more than 10 million employees in 50 countries from about 6,000 organizations varying in size, branch of industry, maturity, and structure. The assessed key factor is trust. “From the employee’s perspective, a great workplace is one where they trust the people they work for, have pride in what they do, and enjoy the people they work with” (Great Place to Work, 2014b, “Great Place to Work Model”). In both categories, Germany’s Best Workplaces 2012 and Best Workplaces in Health Services, data were based on (a) an employee survey measuring quality of organizational culture by focusing on the dimensions fairness, credibility, respect, pride, and camaraderie and (b) an evaluation of organization-specific measures in HR management such as leadership or career development (“Germany’s Best Workplaces 2012”, 2012, para. 8; “Best Workplaces in Health Services”, 2012, para. 4). The ranking of the *Trendence Institute* (2013) subsumed companies that were rated on brand recognition value and the companies’ attractiveness by young professionals with a maximum of eight-year work experience (Young Professional Barometer) and final-year university students (Trendence Graduate Barometer) in the categories Business, Engineering, Information Technology, and Law as the most attractive employers in Germany.

Data were collected cross-sectionally in two waves with a standardized online questionnaire created with SoSci Survey (Leiner, 2013). In the first wave (December 2012 until February 2013), T&D professionals (*N* = 278) of the companies that have been awarded as “best employers in Germany” ranked by the Top Employers Institute and Great Place to Work were contacted first by phone and second via email. In the second wave (March until May 2013), T&D professionals (*N* = 251) of the companies awarded by the Trendence Institute were contacted. Each company participated only once in the study, even though they were repeatedly represented in several rankings. Participants were informed about the objectives of the research study and their benefits of participating in the study in such a manner, for example, as to receiving a benchmark as well as a checklist of evidence-based recommendations and best practices for maximizing training effectiveness free of charge. The final sample that was considered in the analyses comprised 134 companies that fully completed the questionnaire; thus, a response rate of 27.5% was achieved (recruited companies interested in the study and receiving an email invitation were *N* = 487). To give an informative description of the participants of the study, demographical data were collected about the companies (i.e., branch of industry, total number of employees, and turnover) and about T&D professionals (i.e., number of years responsible for training and job position).

Characteristics of companies

Branches of industry. Overall, companies from more than 14 different branches of industry participated in the study (see Figure 7). Services industries (14.2%) as well as health care and social services (13.3%) were most frequently represented. The category labeled "others" (6.7%) comprised the companies that could not be assigned to any explicit branch of industry (e.g., defence, development cooperation).

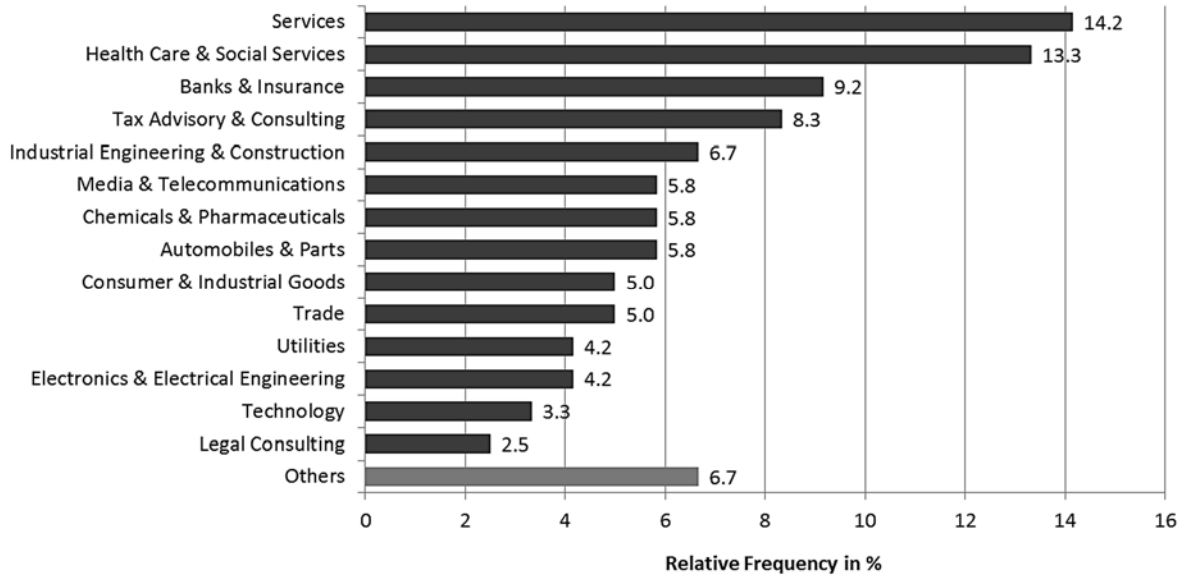

Figure 7. Relative frequencies of the companies by branches of industry adapted from the Industry Classification Benchmark (FTSE International Limited, 2014, "Industry Classification Benchmark"). *n* = 120.

Total number of employees and turnover. The size of enterprise was determined by the number of employees and turnover. According to the recommendation of the European Commission (IfM-Institut für Mittelstandsforschung Bonn, 2014), companies are defined as small and medium-sized enterprises when less than 250 workers are employed and the turnover per fiscal year does not exceed 50 million Euro. Focusing on the national outreach of this study, the number of employees in Germany is presented in Figure 8. Mostly, large-scale enterprises participated in the study (85.0%), whereas only 15.0% of the companies were among the small and medium-sized enterprises.

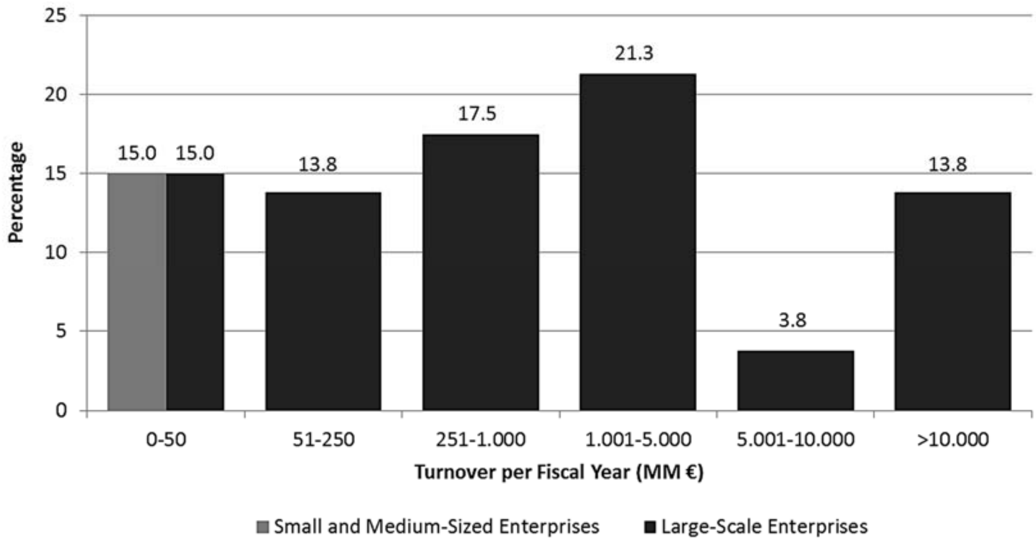

Figure 8. Total number of employees as a function of turnover per fiscal year (MM €). Small and medium-sized enterprises employ less than 250 workers and the turnover per fiscal year does not exceed 50 million Euro (IfM-Institut für Mittelstandsforschung Bonn, 2014). n = 80.

Characteristics of T&D professionals

The questionnaire was largely completed by T&D professionals (83.5%) as displayed in Figure 9. 7.8% were project managers or assistant managers. In 4.3% of the cases, the chief executive officer participated in the study. Participants had possessed a great wealth of experience: six years of training responsibility on average; ranging from less than one year up to 30 years (see Figure 10). More than half of them (52.1%) had been responsible for training for more than three years (M = 5.95, SD = 6.17).

Figure 9. Percentage of T&D professionals as a function of positions. n = 115.

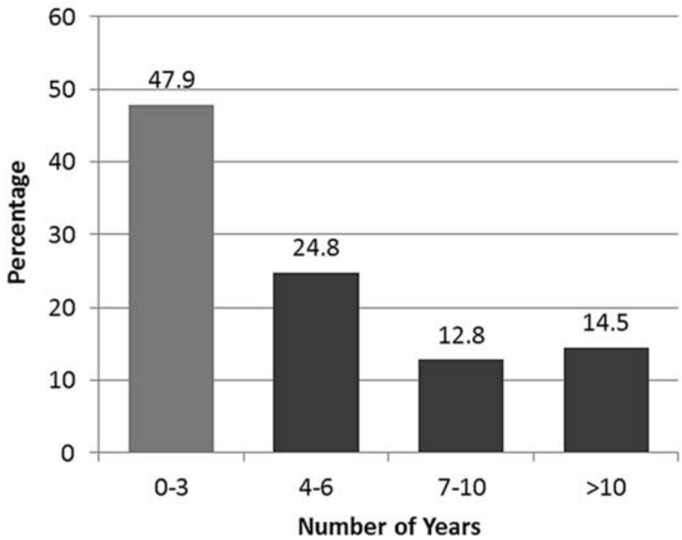

Figure 10. Percentage of T&D professionals as a function of expertise in number of years responsible for training. n = 117.

Instrument

Implementation of e-learning and classroom training. We gathered data about e-learning-specific issues (i.e., general implementation, period of implementation, implementation of various forms, topics, trainees) and classroom-specific issues (i.e., topics, trainees).

Implementation of actions for maximizing training effectiveness. Items for evaluating actions for maximizing training effectiveness were based on the checklist by Salas et al. (2012). For our research study, this checklist served as core approach since it subsumes a wide range of both theoretical models as well as evidence-based recommendations and best practices for maximizing training effectiveness. In addition, the questionnaire was supplemented by items derived from Wirth et al. (2009). We used the translation-back translation technique (e.g., Van de Vijver & Leung, 1997) in order to construct a German-speaking questionnaire. Participants' ratings ranged from 1 (*never*) to 5 (*always*) on a 5-point Likert scale. Items were subsumed in scales.

The questionnaire contained actions for maximizing training effectiveness for e-learning only (9 items) as presented in Table 22 and actions for maximizing training effectiveness for both training settings (see Table 23, Table 24, and Table 25). For the latter, as each item was presented in a parallel structure, the aimed direct comparison was possible (each item had to be answered for both training settings). Here, we surveyed the implementation of actions for maximizing training effectiveness separately for before (21 items), during (19 items), and after (22 items) training.

Table 22. *Scale and Sample Items of E-Learning-Specific Actions for Maximizing Training Effectiveness Adapted from Salas et al. (2012)*

Scale	α	*n*	Sample item
Adequately design e-learning settings	.60	5	We consider carefully which training topics can be delivered via e-learning. We ensure that our e-learning actions are based on sound instructional design. In case of questions and problems, participants get support from a hotline or an IT department. Trainees are provided with feedback. Participants are offered sufficient structure and guidance to allow them to make decisions about their learning experience.
Use simulation appropriately	.91	4	Simulations are used to train complex and dynamic skills (particularly those that may be dangerous). We ensure that the simulation is job relevant, even if it is not identical to the job. When implementing simulations, opportunity for performance diagnosis and feedback is built in. After implementing simulations, practice in the workplace is guided.

Note. α = Cronbach's α for e-learning. *n* = number of items used for each particular scale.

Table 23. *Scales and Sample Items of Actions for Maximizing Training Effectiveness Before Training Adapted from Salas et al. (2012)*

Scale	α EL	α CT	*n*	Sample item
Conduct a job-task analysis	.65	.48	4	We consider conducting a cognitive task analysis for knowledge-based jobs (= method of analysis for the course of cognitive processes).
Conduct a person analysis	.64	.37	4	We gear the structure of our training especially to the needs of elderly employees.
Conduct an organizational analysis	-	-	2	We determine if our company guidelines (culture, norms) support the training.
Schedule training	-	-	2	We schedule training close to when trainees will be able to use on the job what they have learned.
Notify employees	.76	.68	4	We communicate clear expectations about the training.
Establish attendance policies	-	-	2	We use the mandatory label selectively.
Prepare supervisors and leaders	-	-	2	We prepare supervisors to support their employees.
Select trainer	-	-	1	When selecting trainers, educational competence is very important to us.

Note. α = Cronbach's α for e-learning calculated at the minimum of three items. EL = e-learning. CT = classroom training. *n* = number of items used for each particular scale.

Table 24. *Scales and Sample Items of Actions for Maximizing Training Effectiveness During Training Adapted from Salas et al. (2012)*

Scale	α EL	α CT	*n*	Sample item
Build self-efficacy	.79	.64	4	Training is delivered in a way that builds trainees' belief in their ability to learn.
Promote a learning orientation	-	-	1	During training, we encourage trainees to participate actively in training rather than to act passively.
Boost motivation to learn	.81	.82	3	We ensure that the training is perceived as relevant and useful.
Use a valid training strategy and design	.72	.67	4	During training, demonstrations of bad behavior are given.
Build in opportunities for trainees to engage in transfer-appropriate processing	.68	.58	3	Exercises incorporate features that require trainees to engage in the same cognitive processes during training compared to those in the transfer environment.
Promote self-regulation	-	-	2	We keep participants on task by encouraging their self-monitoring.
Incorporate errors into training	-	-	2	During training, participants are encouraged to make errors.

Note. α = Cronbach's α for e-learning calculated at the minimum of three items. EL = e-learning. CT = classroom training. *n* = number of items used for each particular scale.

Table 25. *Scales and Sample Items of Actions for Maximizing Training Effectiveness After Training Adapted from Salas et al. (2012)*

Scale	α EL	α CT	*n*	Sample item
Remove obstacles to transfer	-	-	2	We ensure that participants have ample opportunity to apply what they have learned and receive feedback.
Provide tools and advice for supervisors	-	-	2	We ensure that supervisors are equipped to reinforce trained skills.
Encourage use of real-world debriefs	.85	.73	6	After training, we reinforce lessons learned (e.g., praise by supervisors).
Provide other reinforcement and support mechanisms	.64	.46	3	At work, we provide participants with job aids.
Clearly specify the purpose of evaluation	-	-	2	By evaluating the training, we determine whether we accomplished our goals or purposes.
Consider evaluating training at multiple levels	.84	.79	7	We evaluate results on company level (e.g., increase in productivity).

Note. α = Cronbach's α for e-learning calculated at the minimum of three items. EL = e-learning. CT = classroom training. *n* = number of items used for each particular scale.

Characteristics of companies and T&D professionals. In addition, we investigated characteristics of the companies (e.g., branch of industry, total number of employees, and turnover) and T&D professionals (e.g., job position, number of years responsible for training).

5.1.3 Results

Implementation of e-learning and classroom training

General implementation of e-learning. More than two-thirds of the companies surveyed offered e-learning as part of their HRD activities, namely banks and insurance or services (see Figure 11). Only one third of the companies did not implement e-learning for knowledge transfer in their training (e.g., health care and social services).

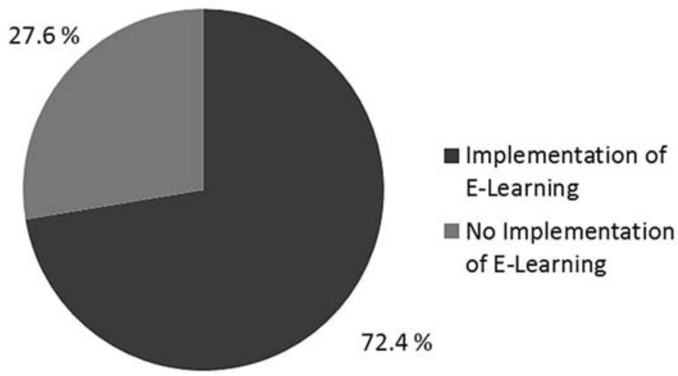

Figure 11. Percentage of the companies reporting general implementation of e-learning. *N* = 134.

Period of implementation of e-learning. When asking for the period of time for which e-learning has been implemented as part of training settings, a mean of eight years with a high standard deviation was found (M = 7.82, SD = 4.82, [1, 25]) as displayed in Figure 12. More than half (53.3%) of the companies with a given answer had been using e-learning for more than six years, namely banks and insurances as well as services. These companies had a significantly higher number of employees (in Germany with r = .25, p < .10 and internationally with r = .30, p < .01) and higher turnover with r = .32, p < .01. Only 23.3% of the companies had been using e-learning for 0 to 3 years.

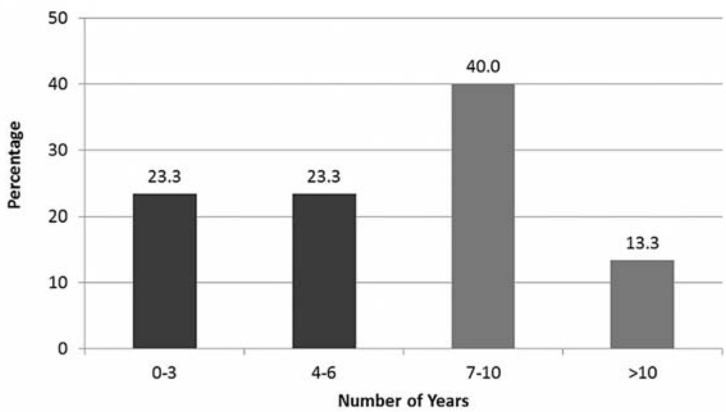

Figure 12. Percentage of the companies implementing e-learning in number of years. *n* = 60.

Implementation of various forms of e-learning. For each form of e-learning, a short definition was provided. Web-based training (49.5%), learning platforms (36.1%), blended learning (27.8%), and computer-based training (26.8%) were most commonly used, as highlighted in dark blue in Figure 13 representing only *fairly common* or *very common* use of various e-learning forms. Wikis and blogs (20.6%), podcasts (15.5%), and virtual classrooms (14.4%) were comparatively rarely implemented, as highlighted in blue. Business games and simulations have not frequently used in most of the companies so far, as highlighted in light blue.

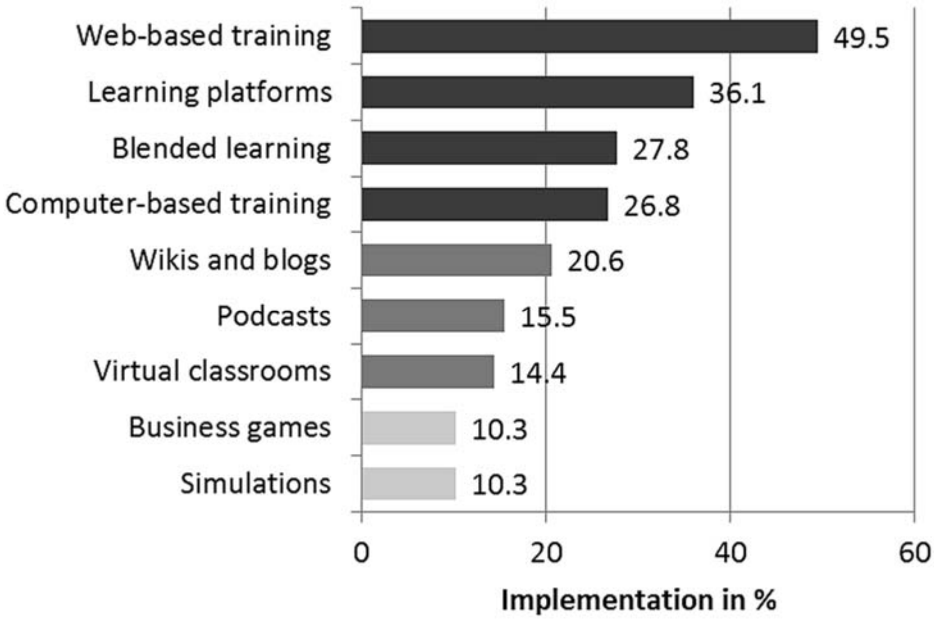

Figure 13. Implementation of various forms of e-learning in percentage of the companies reporting *fairly common* or *very common* use. n = 97.

Topics in e-learning and classroom training. With regard to topics taught we found that trainings imparting declarative (factual) knowledge such as hard skills (e.g., software and legal regulation trainings) were significantly more frequently chosen and thus trained in e-learning settings than in classroom training ($p < .01$). At the same time, soft skills (e.g., foreign languages) were more frequently trained in classroom training ($p < .001$). Thus, Hypothesis 1 was confirmed (Figure 14). Analyses showed that no training setting dominated the other with respect to product training, professional expertise, and methodological competence (which can refer to training of both hard and soft skills) (all ps ≥ .05).

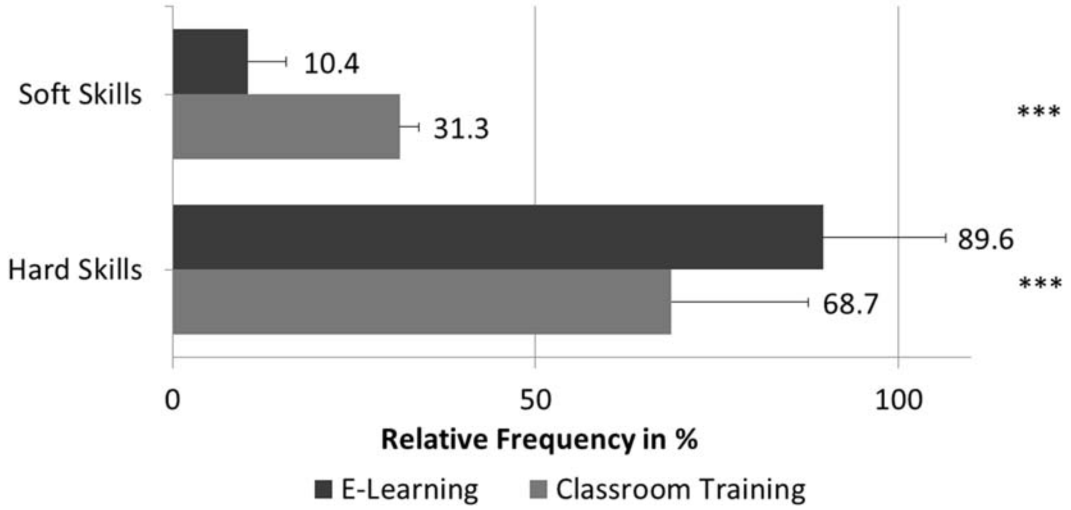

Figure 14. Percentages of general use of topics in e-learning (n = 91) and classroom training (n = 115). * $p < .05$ with Cohen's $d < 0.5$ (small). ** $p < .01$ with Cohen's $0.5 \leq d < 0.8$ (medium). *** $p < .001$ with Cohen's $d \geq 0.8$ (large) according to Cohen (1988).

Trainees in e-learning and classroom training. In both training settings, the target group comprised clerks, managers, skilled employees, and sales staff. Managers were slightly more likely to be trained in classroom training (p = .090). Otherwise, no training setting dominated the other. Due to a reduced number of employees within the company's staff, consultants, apprentices, and casual laborers were not primarily focused on in any specific training setting.

Implementation of actions for maximizing training effectiveness

Implementation of e-learning-specific actions for maximizing training effectiveness. The majority of the companies surveyed *almost always designed their e-learning settings adequately* (scale level: *M* = 4.06, *SD* = 0.60) as dark blue-highlighted single items display in Figure 15. To be more specific, data on the single-item level revealed that the companies *almost always ensure that their e-learning actions are based on sound instructional design* (*M* = 4.27, *SD* = 0.81), and that they *consider carefully which training topics can be delivered via e-learning* (*M* = 4.25, *SD* = 0.90). Further, *in case of questions and problems, participants get support from a hotline or an IT department* (*M* = 4.16, *SD* = 1.03)*, trainees are provided with feedback* (*M* = 3.82, *SD* = 1.07)*, and they are offered sufficient structure and guidance to allow them to make decisions about their learning experience* (*M* = 3.83, *SD* = 0.87). Answering our research question as to what extent companies implement e-learning-specific actions for maximizing training effectiveness that maximize the success of learning and success of transfer, we concluded that the companies were good at maximizing success of transfer as indicated by high values in designing their e-learning settings adequately even though there is still some room for improvement.

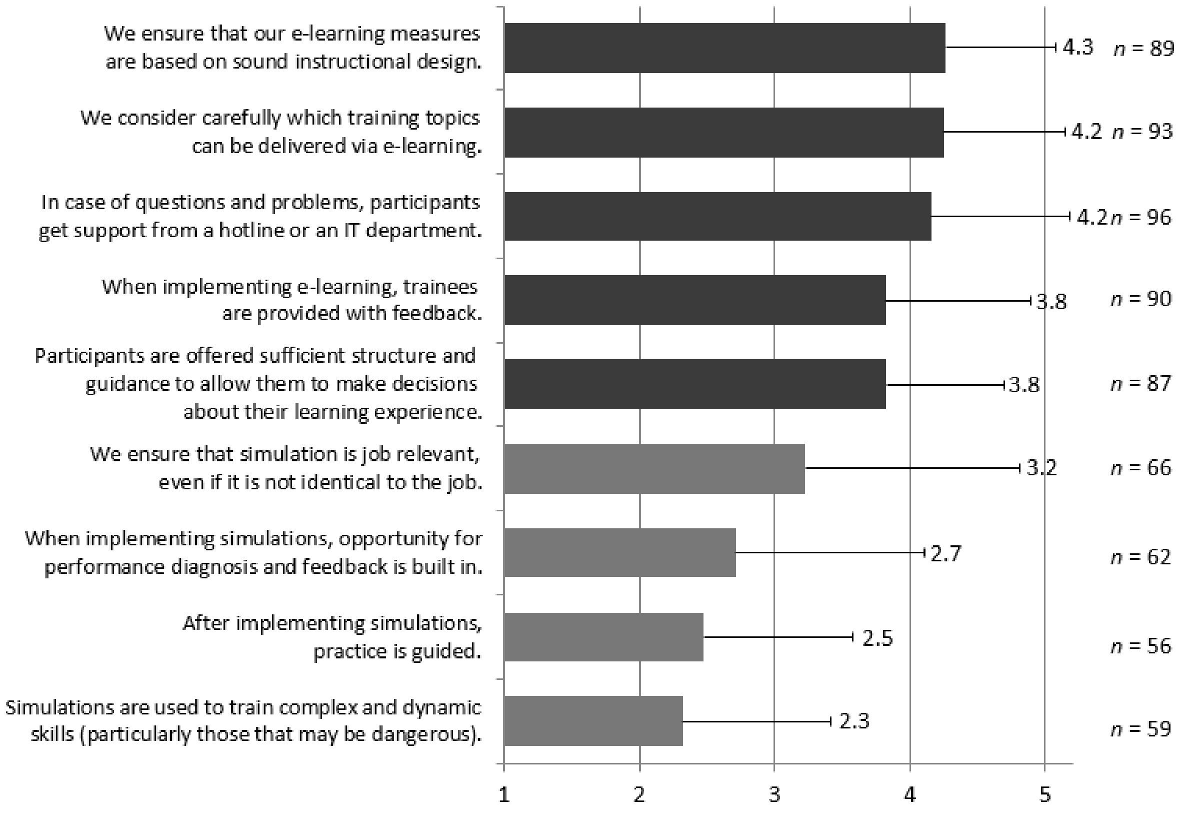

Figure 15. Implementation of e-learning-specific actions for maximizing training effectiveness. Mean on Likert scale ranging from 1 (*never*) to 5 (*always*). *n* varies between 56 and 96.

However regarding the scale measuring the *appropriate use of simulations*, a relatively low value with high variance and lots of missing data was obtained as displayed in light blue in Figure 15 (M = 2.62, SD = 1.18). Here, the companies showed low values (*almost never*) in *using simulations to train complex and dynamic skills* (M = 2.32, SD = 1.10) and also in *guiding practice* (M = 2.47, SD = 1.10).

Implementation of actions for maximizing training effectiveness before training. Descriptive analyses revealed high means for both forms of learning, all Ms ≥ 3.33 (see Figure 16). Thus, it was *sometimes* to *almost always* the case for both training settings that actions for maximizing training effectiveness were implemented before training. Comparing both training settings, all investigated actions for maximizing training effectiveness were more often applied in classroom training. Notably, actions in e-learning showed higher standard deviations and more missing data than in classroom training. T-Tests showed significant mean differences in favor of classroom training in almost all scales, all ps ≤ .01: *conduct a job-task analysis*, *conduct a person analysis*, *conduct an organizational analysis*, *notify employees*, and *prepare supervisors and leaders*. No significant differences between training settings occurred for *schedule training* and *establish attendance policies.*

Looking closely at single items, results revealed a demand for adapting the training environment to meet the needs of older workers (e-learning: M = 2.09, SD = 1.13; classroom training: M = 2.30, SD = 1.04). In addition, *conducting a cognitive task analysis for knowledge-based jobs* was *almost never* considered (e-learning: M = 1.67, SD = 0.93; classroom training: M = 1.73, SD = 0.97). The mandatory label was *sometimes* used selectively (e-learning: M = 3.28, SD = 1.49; classroom training: M = 3.20, SD = 1.54), thus leaving room for voluntary attendance that helps to ensure the learner’s motivation and course attendance, and that helps to maximize the success of transfer (for meta-analytical details, see Blume et al., 2010). Especially in e-learning settings, the companies only *sometimes examined teamwork demands, if needed* (e-learning: M = 2.93, SD = 1.38). Since the single item regarding the *selection and competence of the trainer* was phrased rather classroom-specific, no further comparative analysis was conducted.

In summary, results confirmed Hypothesis 2a such that fewer actions for maximizing training effectiveness before training are implemented in e-learning compared to classroom training settings. Even though high means in both training settings were obtained, there is room for potential improvement especially in meeting the needs of older workers.

Figure 16. Implementation of actions for maximizing training effectiveness before training. Mean on Likert scale ranging from 1 (*never*) to 5 (*always*). *n* varies between 87 and 92. * $p < .05$ with Cohen's $d < 0.5$ (small). ** $p < .01$ with Cohen's $0.5 \leq d < 0.8$ (medium). *** $p < .001$ with Cohen's $d \geq 0.8$ (large) according to Cohen (1988).

Implementation of actions for maximizing training effectiveness during training. Results during training revealed the same underlying structure as before training. That is, (a) actions to maximize training effectiveness were more often implemented in classroom training than in e-learning settings (all *p*s < .01), confirming Hypothesis 2b; (b) standard deviations were greater in e-learning; and (c) more data were missing in e-learning (see Figure 17). Compared to results before training, we found even higher means in both training settings (e-learning: $M_{\text{min to max}}$ = 3.4 to 4.1, *sometimes* to *almost always*; classroom training: $M_{\text{min to max}}$ = 3.8 to 4.5, *almost always* to *always*).

Looking closely at single items, we noticed in both training settings that trainees were *only sometimes encouraged to make mistakes during training* (e-learning: M = 2.90, SD = 1.10; classroom training: M = 3.39, SD = 0.92) and *demonstrations of bad behavior* were too rarely provided (e-learning: M = 3.30, SD = 1.21; classroom training: M = 3.58, SD = 1.04). In addition, especially in e-learning, a relatively low value was obtained for *giving meaningful and diagnostic feedback* to the trainees (e-learning: M = 3.19, SD = 1.06; classroom training: M = 3.78, SD = 0.82).

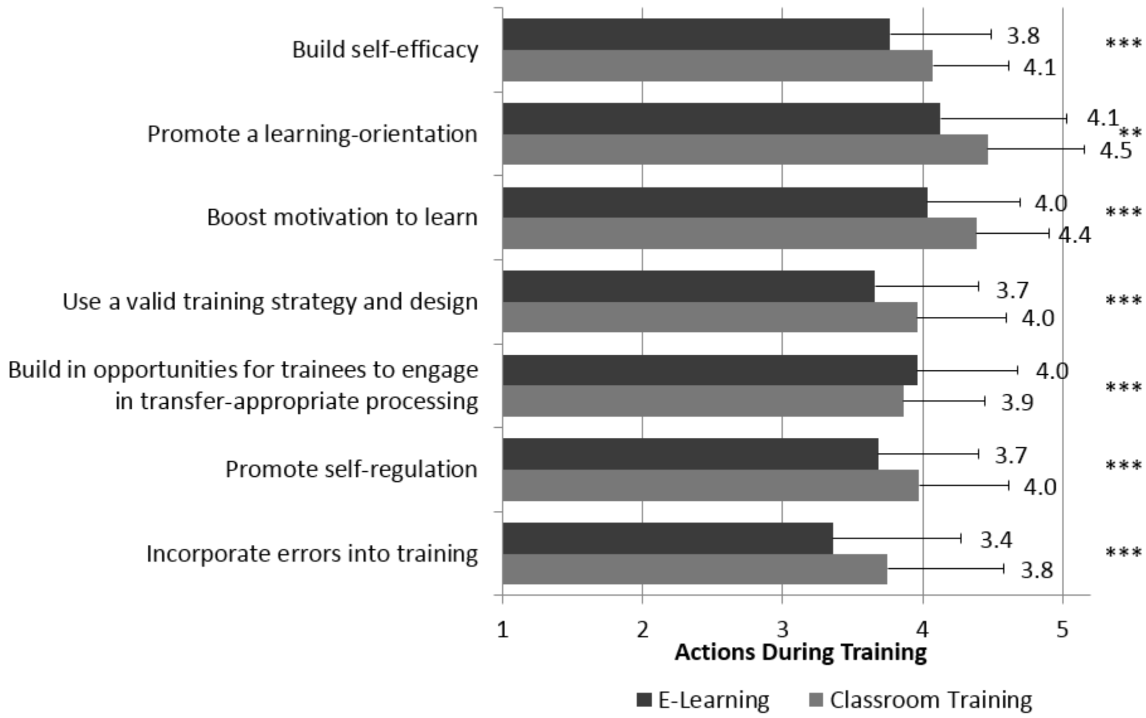

Figure 17. Implementation of actions for maximizing training effectiveness during training. Mean on Likert scale ranging from 1 (*never*) to 5 (*always*). *n* varies between 68 and 84. * $p < .05$ with Cohen's $d < 0.5$ (small). ** $p < .01$ with Cohen's $0.5 \leq d < 0.8$ (medium). *** $p < .001$ with Cohen's $d \geq 0.8$ (large) according to Cohen (1988).

Implementation of actions for maximizing training effectiveness after training. In both training settings, means of actions for maximizing training effectiveness after training were comparatively lower than means before and during training. Again, in e-learning settings, data varied more in terms of standard deviation and missing data. Figure 18 displays significant mean differences of actions for maximizing training effectiveness after training in favor of classroom training in all scales (all *p*s < .05), confirming Hypothesis 2c.

The scale *clearly specifying the purpose of evaluation* scored highest in both training settings. Thus, the companies *almost always* determined what they hope to accomplish by evaluating training and they linked all subsequent decisions back to their purpose. Significant differences ($p < .05$) indicating a small effect between both training settings were found for *providing other reinforcement and support mechanisms* (e-learning: $M = 3.10$, $SD = 0.90$; classroom training: $M = 3.20$, $SD = 0.83$) and for *considering evaluating training at multiple levels* (e-learning: $M = 3.11$, $SD = 0.97$; classroom training: $M = 3.26$, $SD = 0.84$). As these scales were relatively low in value compared to others (only *sometimes* implemented), a tremendous potential for improvement is indicated.

Analyses on the single-item level revealed that virtual or real “communities of practice” were *almost never* used to reinforce and support the content trainees learned in training (e-learning: M = 2.39, SD = 1.13; classroom training: M = 2.49, SD = 1.05). Also, *implementing a plan of action (agreements) between the trainee and supervisor* was relatively rare (e-learning: M = 2.46, SD = 1.11; classroom training: M = 2.79, SD = 1.04).

Especially in terms of *considering evaluating training at multiple levels*, we found that the companies surveyed obtained quite high values when measuring participants’ satisfaction with training and success of learning (see Figure 19). However, all other levels dealing with the measurement of success of transfer, training success at the company level, and the use of precise affective, cognitive, and/or behavioral indicators to measure the intended learning outcomes were disregarded (*almost never*). Illustrating this potential improvement, mean values scored on each individual question of the scale *considering evaluating training at multiple levels* are displayed.

Figure 18. Implementation of actions for maximizing training effectiveness after training. Mean on Likert scale ranging from 1 (*never*) to 5 (*always*). n varies between 73 and 81. * p < .05 with Cohen’s d < 0.5 (small). ** p < .01 with Cohen’s 0.5 ≤ d < 0.8 (medium). *** p < .001 with Cohen’s d ≥ 0.8 (large) according to Cohen (1988).

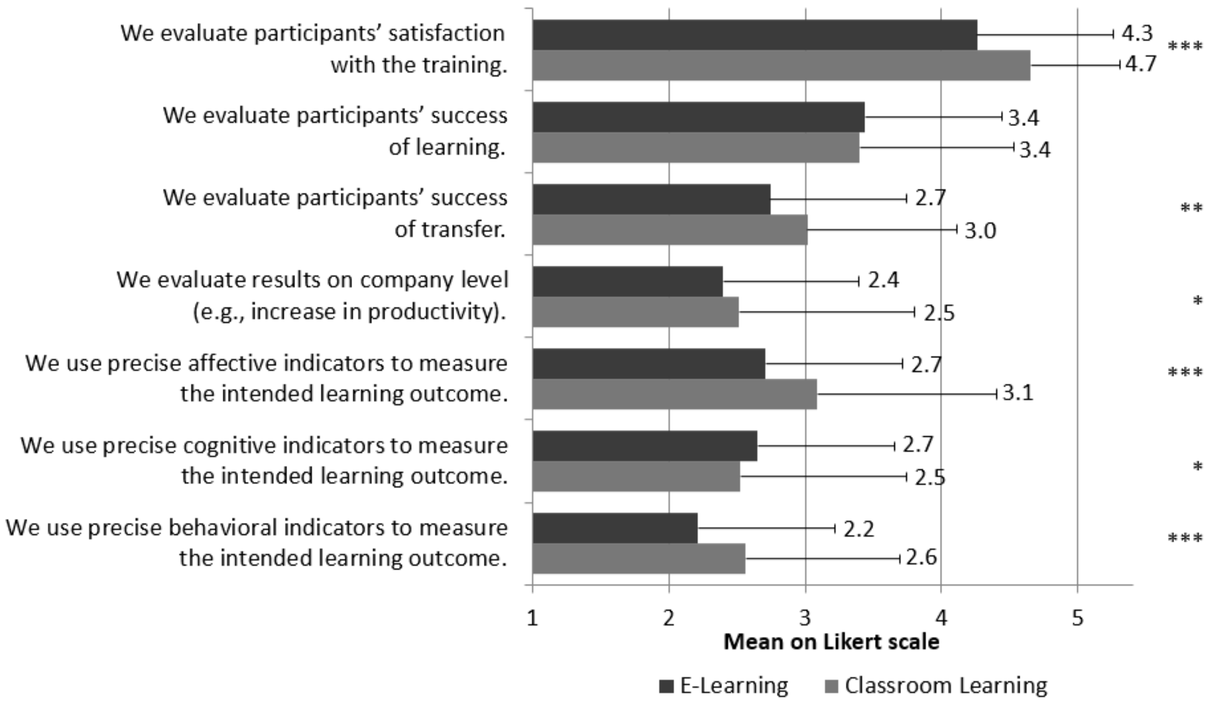

Figure 19. Items of the scale *consider evaluating training at multiple levels* (Salas et al., 2012). Mean on Likert scale ranging from 1 (*never*) to 5 (*always*). *n* varies between 60 and 81. * $p < .05$ with Cohen's $d < 0.5$ (small). ** $p < .01$ with Cohen's $0.5 \leq d < 0.8$ (medium). *** $p < .001$ with Cohen's $d \geq 0.8$ (large) according to Cohen (1988).

Comparison of training settings and time periods. An analysis of variance with two within-subjects factors was conducted to assess the impact of the two different training settings (e-learning, classroom training) on participants' scores on a 5-point Likert scale measuring the implementation of actions for maximizing training effectiveness across three time periods of training (before, during, and after training). There was no significant interaction between training settings and time periods, Wilks' Lambda = .95, $F(1, 80) = 2.11$, $p = .13$, $\eta^2_{partial} = .050$. The substantial main effect comparing the two training settings was significant, Wilks' Lambda = .58, $F(1, 80) = 57.26$, $p < .001$, $\eta^2_{partial} = .42$, suggesting a difference in the implementation of actions for maximizing training effectiveness between the two training settings. Further, there was a substantial main effect for time period, Wilks' Lambda = .38, $F(1, 80) = 63.72$, $p < .001$, $\eta^2_{partial} = .62$, with both training settings showing an overall reduction in scores on the 5-point Likert scale pertaining to the implementation of actions for maximizing training effectiveness across time periods (see Table 26 and Figure 20). Tests of within-subjects contrasts showed a very large effect for differences between before and after training ($p < .001$, $\eta^2_{partial} = .47$) as well as between during and after training ($p < .001$, $\eta^2_{partial} = .61$) with always lower scores for the time period after training. The difference between before and during training revealed a very much smaller effect ($p < .01$, $\eta^2_{partial} = .090$). Thus, Hypothesis 3 was confirmed such that actions for maximizing training effectiveness are least implemented after training compared to before and during training in both training settings.

Table 26. *Implementation of Actions for Maximizing Training Effectiveness Before, During, and After Training in E-Learning and Classroom Training Measured on a 5-Point Likert Scale*

	E-Learning			Classroom Training		
Time period	*n*	*M*	*SD*	*n*	*M*	*SD*
Before training	81	3.60	0.73	81	3.92	0.52
During training	81	3.74	0.58	81	4.07	0.43
After training	81	3.16	0.78	81	3.40	0.63

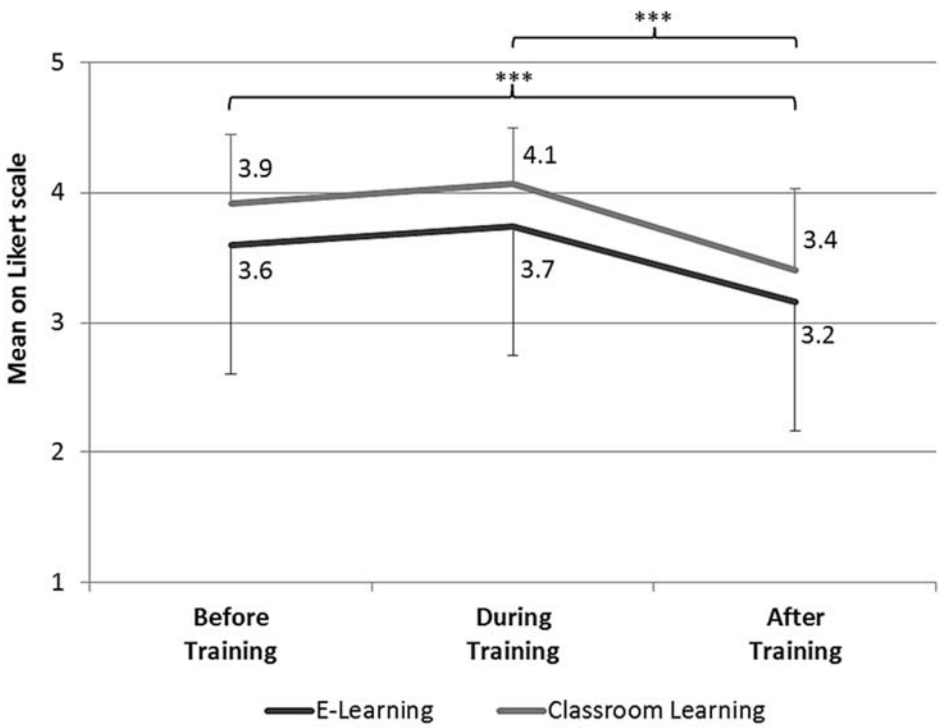

Figure 20. Comparisons of the implementation of actions for maximizing training effectiveness before, during, and after training in e-learning and classroom training. Mean on Likert Scale ranging from 1 (*never*) to 5 (*always*). $n = 81$. According to Cohen (1988), results revealed a large main effect for training setting with $\eta^2_{partial} = .42$ and a large main effect for time period with $\eta^2_{partial} = .62$ but no interaction effect. Tests of within-subjects contrasts showed very large effects for before and after training (*** $p < .001$, $\eta^2_{partial} = .47$) as well as for during and after training (*** $p < .001$, $\eta^2_{partial} = .61$) but only a medium effect for before and during training (** $p < .01$, $\eta^2_{partial} = .090$).

5.1.4 Discussion

Summary with implications for practice

The fourth study aimed at contributing to the specific research gap regarding the application of evidence-based actions for maximizing training effectiveness in corporate practice. It investigated separately for e-learning and classroom training the extent to what evidence-based actions for maximizing training effectiveness are already implemented before, during, and after training in companies that have been awarded as “best employers in Germany”. Before summarizing main findings, the national or even international scope of application of the study as well as general information about implementing e-learning and classroom training is provided at an introductory level.

The sample comprised primarily large-scale enterprises (85.0%) with high turnovers and various branches of industry (especially the branches services, health care and social services, banks and insurance, tax advisory and consulting). These companies reported that their responses to the implementation of actions for maximizing training effectiveness have a national or even international scope of application (74.3%). Due to the selected sample (i.e., large-scale enterprises with high turnovers and great number of employees, awarded as “best employers in Germany”, national or even international scope of application), results can be interpreted mainly nationally and partially internationally. This accounts for a high practical impact for trainees as well as T&D professionals and also demonstrates the relevance of this exploratory research study for research in the field of T&D.

In the following, key findings with their corresponding implications for practice are summarized. In support of the increasing implementation of virtual training settings in corporate practice (MMB, 2010), we identified that more than two-thirds of the companies offered e-learning as part of their HRD activities. This corresponds to an increase of approximately one-third compared to results of an earlier study conducted by the MMB (2010). Experts of the MMB-Trendmonitor assumed that e-learning and blended learning as training settings will play a leading role in the future (MMB, 2012). *Most commonly*, the companies used web-based training (49.5%), learning platforms (36.1%), blended learning (27.8%), and computer-based training (26.8%). More than half of these companies had been using e-learning for more than six years (eight years on average), indicating good experiences with e-learning as a training setting, which accounts for a high accuracy of the obtained measures questioned. As the study identified no preference for any target group of trainees in one of the two training settings, results regarding topics in e-learning and classroom training confirmed that hard skills are more frequently trained in e-learning settings, whereas soft skills are more frequently trained in classroom training (Hypothesis 1).

Responding to the current shift from classroom training to e-learning and answering our research question as to what extent companies implement e-learning-specific actions for maximizing training effectiveness, we concluded that the companies were good at maximizing success of transfer as indicated by high values in designing their e-learning settings adequately, for example, in terms of carefully considering which training topics can be delivered via e-learning ($M_{\text{min to max}}$ = 3.8 to 4.3). Nevertheless, there is still room for improvement especially for providing feedback, structure, and guidance, so that trainees are able to engage in self-regulatory processes and to reflect and adjust their self-paced learning process (Gravill & Compeau, 2008). Feedback loops help trainees to further monitor their progress toward goals and thus they enhance learning (Cannon & Witherspoon, 2005; Hattie & Timperley, 2007; Kluger & DeNisi, 1996; Salas et al., 2012). Contrary to high values previously examined, the companies showed low values (*almost never*) in using simulations appropriately. Here, lots of missing data were obtained indicating that fewer companies invested in simulations in order to train, for example, complex and dynamic skills. Additionally, when asking for the implementation of various e-learning forms, only 10.3% *fairly commonly* or *very commonly* used simulations. Possibly, the companies rarely deal with situations in which simulations are the training setting of first choice, for example, those that involve particularly dangerous actions. Well-designed simulations enhance learning, improve performance, and help minimize errors. It might be the case that this potential of simulations is still untapped (Salas et al., 2012).

As to the main research focus, results confirmed that fewer actions for maximizing training effectiveness are implemented in e-learning compared to classroom training settings across all three time periods–before (Hypothesis 2a), during (Hypothesis 2b), and after training (Hypothesis 2c). As the items for e-learning and classroom training were presented in a parallel structure (each item had to be answered for both training settings), the aimed direct comparison was possible. Even though high standard deviations occurred especially in e-learning settings, on average, there were great differences between both training settings for most of the scales ($p \leq .001$). One of the reasons for this outcome could be that e-learning is constructed and implemented too fast or that not enough expertise is present to incorporate the variety of well-investigated evidence-based actions for maximizing training effectiveness known from the current research literature of T&D (Salas et al., 2012). Another reason could be that the costs for e-learning settings, especially software, are much higher compared to classroom training settings (Bedwell & Salas, 2010; MMB, 2010).

Speaking to the highly relevant transfer problem, our results confirmed that actions for maximizing training effectiveness are least implemented after training ($M_{\text{min to max}}$ = 2.9 to 4.0) compared to before ($M_{\text{min to max}}$ = 3.3 to 4.2) and during training ($M_{\text{min to max}}$ = 3.4 to 4.5) in both e-learning and classroom training (Hypothesis 3). These findings are in line with previous findings regarding the transfer problem (e.g., Baldwin & Ford, 1988; Grossman & Salas, 2011; Saks et al., 2014). More specifically, on the one hand, our results affirmed findings by Van Buren and Erskine (2002) that companies investigated rather in transfer support actions before and during training instead of after training. On the other hand, our results affirmed findings by Salas et al. (2012) that companies tend to focus on aspects during training as indicated by the highest means of actions for maximizing training effectiveness. However, the aspects after training are most relevant to maximizing training success on a long-term basis, for example, support from supervisors and colleagues, positive and negative consequences, phases of exercises, or meaningful and analytical feedback (Hattie & Timperley, 2007; Salas et al., 2012).

Focusing on scales or items with relatively low values, we concluded the following issues and deduced corresponding practical implications. Outstandingly before training, low values were obtained for adapting the training environment to meet the needs of older workers. This finding calls for a demand of ensuring the fit with trainees' needs particularly in terms of an aging process of employees. During training, we found low values for *incorporating errors into training*. As integrating error management during training correlated with success of transfer (d = 0.44), it is highly recommended to equip trainees to deal with situations containing errors on the job (Keith & Frese, 2008). After training, *encouraging the use of real-world debriefs* seems necessary to foster in the future as this is a simple action to maximize the success of transfer (Salas et al., 2012). Consequently, this action promotes adequate mental models, content retention, self-efficacy, and motivation and thus improves job performance (Salas et al., 2012; Smith-Jentsch, Cannon-Bowers, Tannenbaum, & Salas, 2008; Tannenbaum, Smith-Jentsch, & Behson, 1998). We identified a potential of improvement for *providing other reinforcement and support mechanisms* (e-learning: M = 3.10, SD = 0.90; classroom training: M = 3.20, SD = 0.83). Therefore, companies should, for example, provide job aids, access to knowledge repositories, or communities of practice to improve job performance (Gallupe, 2001; Rosenberg, 1995; Salas et al., 2012; Wenger, 1998; Wenger, McDermott, & Snyder, 2002). Furthermore, almost all items of the scale *considering evaluating training at multiple levels* were only *almost never* to *sometimes* implemented after training. Thus, our results confirmed the lack of systematic evaluation after training which was consistently found in other studies (e.g., Patel, 2010). As the importance of monitoring the learning process for success of learning and success of transfer has been widely demonstrated (e.g., Saks & Burke, 2012), we emphasized that more systematic evaluation on multiple levels is needed in corporate practice in the future. For example, based on Kirkpatrick's (1959) evaluation features, measuring (a) participants' success of learning and success of transfer, (b) results on company level, and (c) intended learning outcomes via affective, cognitive, and/or behavioral indicators should be considered (Patel, 2010; Saks & Burke, 2012; Salas et al., 2012). Adequately evaluating trainings allows well-grounded decisions about training including any necessary modifications and to maintain the training's effectiveness (Salas et al., 2012).

In summary, even though high means in both training settings were obtained, primarily before and during training, there is room for potential improvement especially in meeting the needs of older workers, integrating error management, encouraging the use of real-world debriefs, providing reinforcement and support mechanisms, and evaluating training at multiple levels.

Limitations

Content limitations are discussed first followed by methodological limitations. Content limitations refer to the primarily usage of evidence-based actions for maximizing training effectiveness presented by Salas et al. (2012). The following questions still remain open: What specific features of new technologies do really contribute to learning? How and why do specific types of training result in learning? What higher order skills do experts use to execute a task? And what is the contribution of informal learning?

Methodological limitations refer to the selected sample and survey format. With a response rate of 27.5%, 134 of 487 recruited companies interested in the study and receiving an email invitation fully completed the questionnaire; thus, results may not be generalizable across the companies that have been awarded as "best employers in Germany". Data of the companies were based on self-reporting of T&D professionals. We asked the person most responsible for the company's HRD activities to fill in the questionnaire precisely. However, there is no proof that this person represented the head of T&D department. To confirm our results and to gain a higher impact, more companies with a national or even international scope of application and multidimensional assessments should be investigated, for example, by not only including T&D professionals but also trainees, colleagues of trainees, or supervisors.

We used a standardized online questionnaire to benefit from online assessment advantages that are especially no limitations in time and space as well as cost savings (Bortz & Döring, 2006; Hussy, Schreier & Echterhoff, 2013). However, this survey format limited our control of framework conditions when completing the questionnaire compared to an experimental condition. Thus, artefacts are possible regarding whether participants were influenced or even disturbed. Further, the online questionnaire provided no possibility to clarify the meaning of items–even though samples were given for some but not all items that might be awkwardly phrased, vague, or inarticulate (Hussy et al., 2013; Moosbrugger & Kelava, 2012). As the wording of some items sounded somewhat psychological or complex or notional–especially for non-psychologists–it might have been difficult to clearly understand the meaning of the items when no sample was given in parentheses to remove ambiguity. This effect might have caused fewer conscientious answers or missing values especially for the evaluation scale after training. Therefore, samples in the field of corporate practice need to be added in the future. Generally, we assume the effect of regression to the mean or rather a tendency to answer socially desirable for actions for maximizing training effectiveness items as it displays a positive self-signaling of the companies implying self-presentation, also called impression management. According to Schlenker (1980, 2003), it is defined as "people's effort to portray themselves in particular ways to others and to make others view them as having certain traits or properties" (Baumeister & Finkel, 2010, p. 154).

For future research implications, it seems interesting to investigate this study longitudinally (Bühner, 2011; Fahrmeir, Künstler, Pigeot, & Tutz, 2011) in order to see how the companies react to the benchmark as well as to the checklist of evidence-based recommendations and best practices for maximizing training effectiveness feedbacked.

5.2 Differences in training success in corporate e-learning and classroom training

5.2.1 Aims and hypotheses

Aims

E-learning settings are increasingly used in the corporate context, and effectiveness studies examining predictors of training success are claimed in the research discussion on training evaluation (Aguinis & Kraiger, 2009). Referring to this, the study aimed at investigating differences in training success across corporate e-learning and classroom training by focusing on the perspective of trainees participating in either one of the two training settings. The goal of this study is to systematically compare objective and subjective training success measures in both training settings. Therefore, the study contributes to experimental training effectiveness studies as specifically in e-learning few effectiveness studies have been conducted. Further, the study provides answers to the question whether training setting-specific patterns of training success change over time. Results of the study provide information about similarities and dissimilarities between corporate e-learning and classroom training and about what type of knowledge–factual or applied–is best achieved in each training setting.

Hypotheses

Objective training success

Comparison of training settings. As performance levels depend on the type of content learned, we investigated performance separately for factual knowledge and applied knowledge. Research studies demonstrated that e-learning dominates the area of hard skills or factual knowledge, thus referring to declarative knowledge (MMB, 2010). Particularly, web-based instruction was somewhat more effective than classroom instruction when teaching declarative knowledge, especially in long courses (Sitzmann et al., 2006). On the contrary, classroom instruction was rather implemented when teaching soft skills or applied knowledge, thus referring to procedural knowledge (MMB, 2010). Nonetheless, both training settings did not differ significantly regarding their effectiveness (Sitzmann et al., 2006).

Assuming stability of these effects across time, factual and applied knowledge were assessed immediately after training (Time 1) and after an interval of six to eight weeks (Time 2), and overall, this leads to the following hypotheses:

Hypothesis 1: *Factual knowledge is more effectively trained in e-learning compared to classroom training settings immediately after training in Time 1 (Hypothesis 1a) and six to eight weeks later in Time 2 (Hypothesis 1b).*

Hypothesis 2: *Applied knowledge is similar effectively trained in e-learning as well as in classroom training settings immediately after training in Time 1 (Hypothesis 2a) and six to eight weeks later in Time 2 (Hypothesis 2b).*

Subjective training success

Comparison of training settings. According to Kirkpatrick's (1959) evaluation model measuring training effectiveness or training success (as it is often referred to in the literature), four levels need to be covered: *reaction*, *learning*, *behavior*, and *organizational results*. Empirical evidence investigating a direct comparison of e-learning and classroom training assessing these levels is rare. Derived from Study 4, we found that actions for maximizing training effectiveness are lower in e-learning compared to classroom training settings; these actions included measures covering Kirkpatrick's scales. Two issues were considered to construct the following hypotheses: First, meta-analytical results by Alliger et al. (1997) demonstrated that affective *reactions* (i.e., satisfaction as part of reactions; Level 1) did not qualify for predicting outcomes of *learning* (Level 2; r = .02), transfer of learning (Level 3; r = .07), or *organizational results* (Level 4; Tharenou et al., 2007). Second, as *organizational results* comprise, for example, reduced costs, turnover, improved quality, increased production, or return on investment (Bates & Khasawneh, 2005), it seems difficult for trainees to assess these constructs (G. G. Wang & Wilcox, 2006). Further, *organizational results* often depend on other extraneous factors that are not related to the training itself (e.g., cyclical economic trend or number of competitors according to Goldstein and Ford [2002]) which can lead to confounding effects. Therefore, we focused on the short-term evaluation outcome learning (scale: *knowledge* referring to success of learning,) and long-term evaluation outcome behavior (scale: *application to practice* referring to success of transfer). Assuming stability of the training setting effect across time and expecting an advantage of classroom training derived from Study 4, this leads to the following hypotheses:

Hypothesis 3: *Success of learning perceived by trainees participating in classroom training is higher compared to e-learning training immediately after training in Time 1 (Hypothesis 3a) and six to eight weeks later in Time 2 (Hypothesis 3b).*

Hypothesis 4: *Success of transfer perceived by trainees participating in classroom training is higher compared to e-learning training immediately after training in Time 1 (Hypothesis 4a) and six to eight weeks later in Time 2 (Hypothesis 4b).*

Comparison of time periods. Often, companies observe that the knowledge and skills acquired in training are insufficiently transferred to the workplace (e.g., Grossman & Salas, 2011) which can be due to few opportunities to perform, little support from supervisors and/or colleagues, or pressure of work (Solga, 2011a). This phenomenon called transfer problem has been confirmed by several scientific studies including meta-analyses (Grossman & Salas, 2011; Saks et al., 2014). According to Saks (2002), 38% of trainees do not transfer immediately after training and this rises up to 56% after six months and up to 66% after one year; thus, success of learning and success of transfer are reduced across time.

Hypothesis 5: *Success of learning and success of transfer perceived by trainees reduce across time in e-learning (Hypothesis 5a) as well as in classroom training (Hypothesis 5b).*

5.2.2 Methodology

Participants and Procedure

Investigating the workplace, this experimental study was conducted in a global family company (large-scale enterprise) headquartered in Germany. The family company is a global manufacturer and distributor of endoscopes, medical instruments, and devices. The study involved participants who were either enrolled in vocational training (vocational track: 54.7%) or a university program (business administration) of the company (academic track: 45.3%). Data was included in the total sample, if the participant filled in the performance test and the questionnaire about training success completely across both time periods. The final sample comprised 86 trainees (27 women, 52 men; 7 trainees did not report gender) at early adulthood (30 were 18 to 21 years old, 49 were above 21 years old; 7 trainees did not report their age).

Regarding the procedure as presented in Figure 21, participants were randomly assigned to the two training setting experimental groups: e-learning (n = 41) or classroom training (n = 45). Participants of the e-learning group had one week to complete the e-learning course about anatomy and pathology of anesthesia and emergency medicine. T&D professionals as experts checked that the e-learning course fulfilled requirements of a good transfer-supportive e-learning program according to Salas et al. (2012). Participants of the classroom training group were trained in the same topic but face-to-face instructed by a trainer for one day. In both training settings, trainees' training success was assessed immediately after completing either the e-learning course or the classroom training (Time 1) and again six to eight weeks later (Time 2). This time period was greater than the four-weeks minimum for adequate transfer processes according to Gnefkow (2008). Training success was measured objectively in terms of a performance test on the training topic as well as subjectively in terms of trainees' self-reporting on various training success scales at each time period. This field experiment with a time-lag design allows for measuring the differences across two time periods (immediately after training and six to eight weeks follow-up).

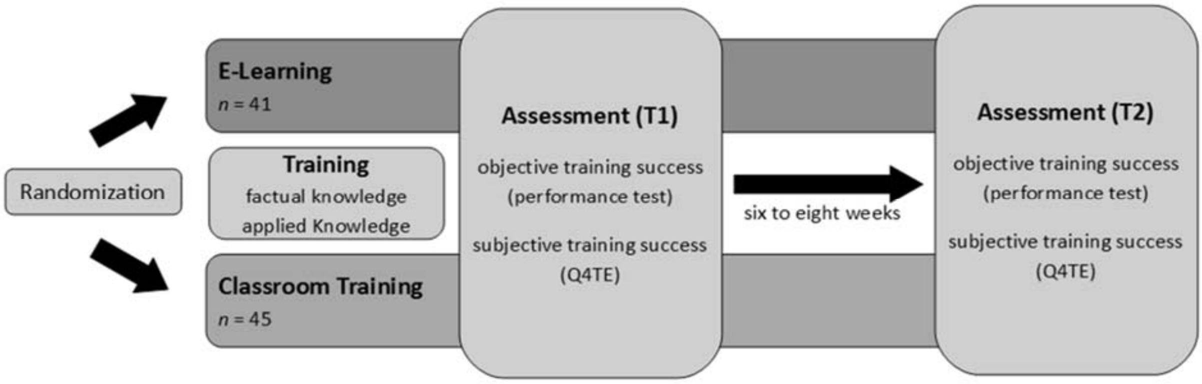

Figure 21. Overview of the study design.

Instruments

Objective training success measures. The performance test was developed by the companies' T&D professionals and experts in the training topic. The test on anatomy and pathology of anesthesia and emergency medicine comprised 23 multiple-choice questions about factual (n = 17, 57 points) and applied knowledge (n = 6, 20 points) resulting in a maximum total test score of 77 points.

Subjective training success measures. To assess or examine training success, Kauffeld and colleagues have produced a self-report measure that is time-efficient, psychometrically sound, and widely applicable across different training contents and training settings (Grohmann & Kauffeld, 2013). It is called Questionnaire for Professional Training Evaluation (initial Q4TE: Kauffeld et al., 2009; Q4TE: Grohmann & Kauffeld, 2013). As this instrument is based on the four-level evaluation model (Kirkpartick, 1959), the Q4TE scales cover all four levels of Kirkpatrick's framework: *satisfaction, utility* (Level 1: reaction), *knowledge*, *self-efficacy* (Level 2: learning), *application to practice* (Level 3: behavior), *individual and global organizational results* (Level 4: results). Specifically, success of learning is measured by the scale *knowledge* and success of transfer is measured by the scale *application to practice*. Our questionnaire consisted of 22 items that were rated on an 11-point response scale ranging from 0% (*completely disagree*) to 100% (*completely agree*) with single steps of 10% increase each (see Table 27).

Table 27. *Scales and Sample Items of Training Success Adapted from the Initial Q4TE (Kauffeld et al., 2009) and Q4TE (Grohmann & Kauffeld, 2013)*

Level	Scale	*n*	α	Sample item
Reaction	Satisfaction	6	.91	I enjoyed the training very much.
	Utility	2	.96	Participation in the training was very useful for my job.
Learning	Knowledge	5	.91	After the training, I know much more about the training contents than before.
	Self-efficacy	2	.84*	After the training, I face occupational difficulties calmlier because I can better rely on my abilities.
Behavior	Application to practice	2	.90	I am very successful in applying what I have learnt during the training in my daily work.
Results	Individual and global organizational results	5	.91*	The application of the training contents has facilitated the work flow in my company.

Note. *n* = number of items used for each Q4TE scale. α = Cronbach's α (Grohmann & Kauffeld, 2013). * Cronbach's α (Kauffeld et al., 2009).

Control variables for training success. In addition, we investigated the Big Five personality traits (10-Item Big Five Inventory; Rammstedt & John, 2007; Barrick & Mount, 1991), proactive personality (Dura & Solga, 2010), transfer factors using the German version of the LTSI (LTSI; Holton et al., 2000; GLTSI; Kauffeld et al., 2008), and demographic characteristics of trainees (e.g., gender, age) as further potential influencing factors promoting training success (e.g., Colquitt et al., 2000). In order to identify different behavior patterns of trainees during assessments, in the second assessment, we additionally measured using a Likert scale, (a) how often trainees talked about the training contents with colleagues, (b) how often trainees repeated the training contents, (c) how often trainees applied the context learned in training, and (d) how useful the training was for trainees' work.

5.2.3 Results

First, preliminary analyses were performed including descriptive statistics and correlations of objective and subjective training success measures separately for both training settings (see Table 28, Table 29, and Table 30). Analyzing control variables that are stable across time, both experimental groups did not differ regarding all Big Five personality traits (conscientiousness: $p = .084$, all other ps $\geq .67$), proactive personality ($p = .51$), and transfer factors (all ps $\geq .16$). Referring the latter construct, consequently, the transfer climate was equal across training settings and therefore, factors influencing the transfer did not account for possible training setting effects. The percentage of women and men was equally distributed across both training settings (e-learning: 29% women, 71% men; classroom: 39% women, 61% men); also age was equally distributed across training settings (e-learning: 37% were 18 to 21 years old, 63% were above 21 years old; classroom: 39% were 18 to 21 years old, 61% were above 21 years old). Thus, as these control variables are very stable across time and did not differ across training settings, they were not considered in the following analyses.

Objective training success

Comparison of training settings. Training success was assessed objectively in a performance test measuring factual knowledge and applied knowledge separately for both training settings.

A mixed between-within subjects multivariate analysis of variance was performed to investigate the impact of the two different training settings (e-learning, classroom training) on trainees' scores on an objective performance test across two time periods (immediately after training and six to eight weeks follow-up). Two dependent variables were assessed: factual and applied knowledge. This analysis assessed the question of which training setting was more effective in increasing trainees' performance scores across the two time periods. There was a significant interaction between training setting and time, Wilks' Lambda = .71, $F(2, 83) = 16.62$, $p < .001$, $\eta^2_{partial} = .29$. This result indicated that the impact of one variable (i.e., training setting) is influenced by the level of the second variable (i.e., time). As this large interaction effect was significant, general conclusions as in main effects are usually not appropriate, and main effects have to be interpreted very carefully.

Follow-up univariate analyses identified the locus of differences between e-learning and classroom training depending on the time period (for factual knowledge, see Figure 22; for applied knowledge, see Figure 23).

First, regarding factual knowledge, an independent-samples *t*-test revealed no significant differences between both training settings immediately after training in Time 1; thus Hypothesis 1a is not supported in Time 1 (p = .34, η^2 = .011, small training setting effect). Across time, a paired-samples *t*-test showed that e-learning trainees significantly improved (p = .008, η^2 = .042, small time effect) and classroom training trainees significantly decreased in factual performance scores from Time 1 to Time 2 (p = .004, η^2 = .053, small time effect). As a result and contrary to Time 1, this pattern across time led to significant differences in factual performance scores between both training settings six to eight weeks later in Time 2 (p = .003, η^2 = .10, moderate training setting effect). Therefore, Hypothesis 1b was confirmed indicating that factual knowledge was more effectively trained in e-learning compared to classroom training settings six to eight weeks after the training course.

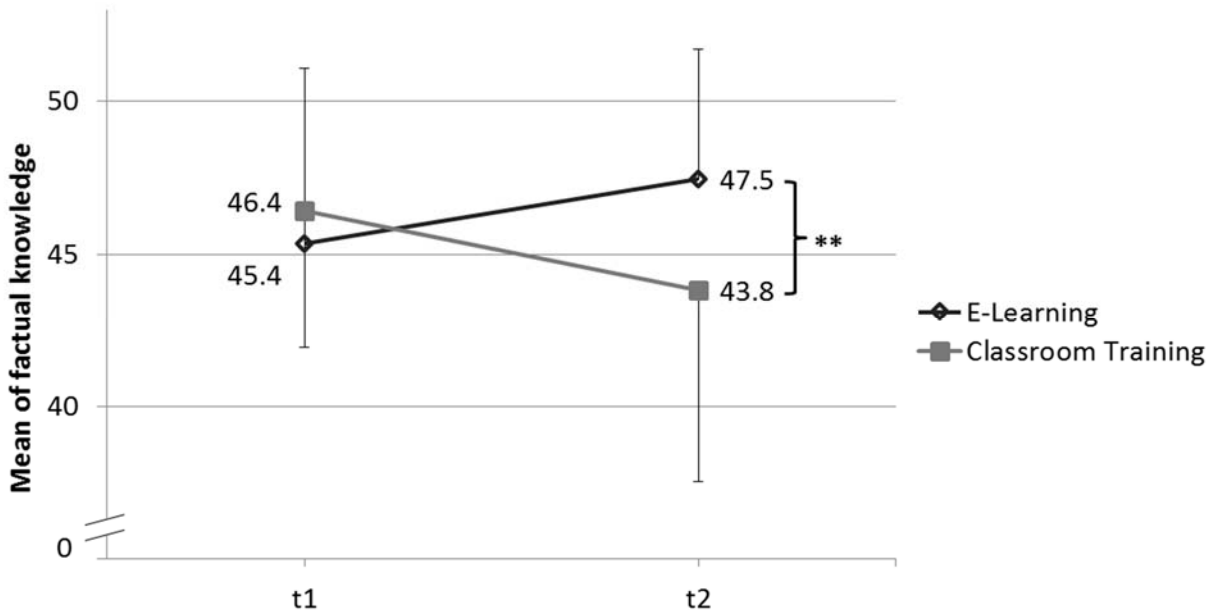

Figure 22. Comparisons of scores obtained in a performance test measuring factual knowledge in e-learning (n = 41) and classroom training (n = 45) immediately after training in Time 1 (p = .34, η^2 = .011, small training setting effect) and six to eight weeks later in Time 2 (** p = .003, η^2 = .10, moderate training setting effect). Means ranged from 0 to 57 points.

Second, regarding applied knowledge immediately after training in Time 1, trainees participating in classroom training scored significantly higher compared to e-learning trainees indicating a large training setting effect (p < .001, η^2 = .29). Thus, Hypothesis 2a was rejected.

Across time, again as for factual knowledge, e-learning trainees significantly improved (p < .001, η^2 = .11, moderate time effect) and classroom training trainees significantly decreased in applied performance scores from Time 1 to Time 2 (p = .001, η^2 = .065, moderate time effect). Again contrary to Time 1, the large training setting effect immediately after training in Time 1 disappeared in support of Hypothesis 2b, assuming that applied knowledge is similar effectively trained in e-learning as well as in classroom training six to eight weeks later in Time 2 (p = .82, η^2 < .001, no training setting effect).

In sum, the two experimental training setting groups showed contrary patterns for each time period; that is, immediately after training in Time 1, applied knowledge scores were higher in classroom training, whereas six to eight weeks later in Time 2, factual knowledge scores were higher in e-learning. These patterns were supported by correlative relationships (see Table 28).

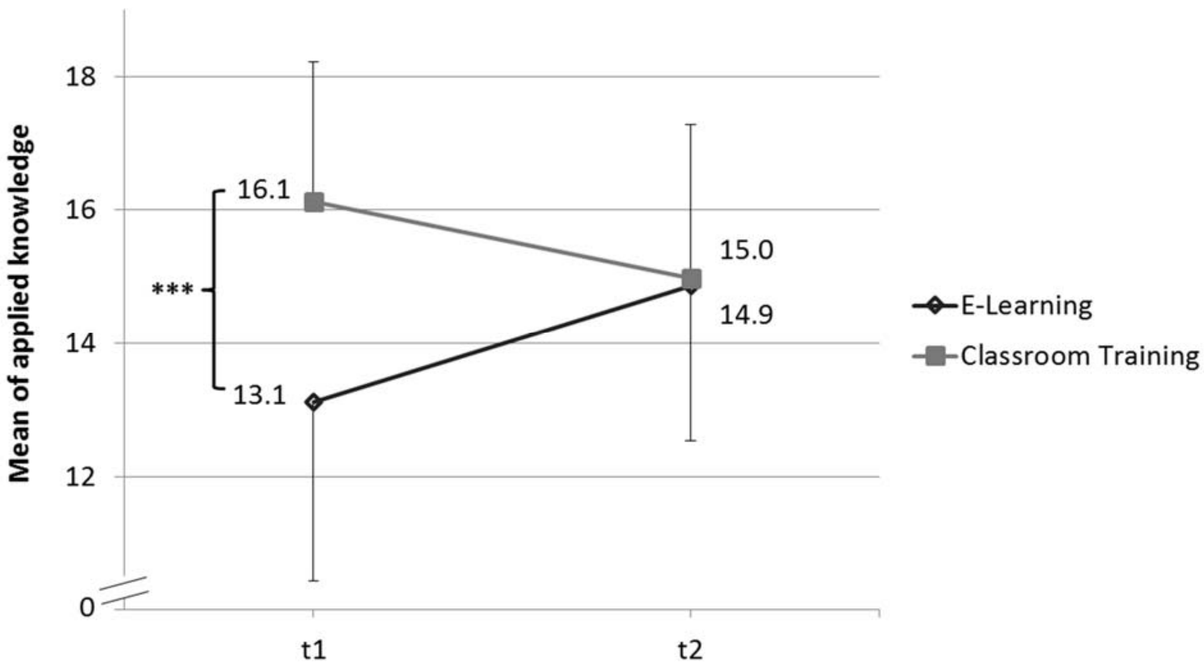

Figure 23. Comparisons of scores obtained in a performance test measuring applied knowledge in e-learning (n = 41) and classroom training (n = 45) immediately after training in Time 1 (*** p < .0001, η^2 = .29, large training setting effect) and six to eight weeks later in Time 2 (p = .82, η^2 < .001, no training setting effect). Means ranged from 0 to 20 points.

Table 28. *Descriptive Statistics and Comparison of Training Settings of Objective and Subjective Training Success Measures Immediately After Training in Time 1 and Six to Eight Weeks Later in Time 2*

			E-Learning (n = 41)		Classroom Training (n = 45)				
Level	Scale	Time	*M*	*SD*	*M*	*SD*	*p*	η^2	*r*
Objective training success									
	Factual knowledge	T1	45.37	5.70	46.44	4.50	.337	.011	.10
		T2	47.46	4.24	43.83	6.29	.003	.103	-.32**
	Applied knowledge	T1	13.12	2.69	16.13	2.09	.000	.287	.54**
		T2	14.86	2.32	14.97	2.29	.818	.001	.02
Subjective training success									
Reaction	Satisfaction	T1	70.04	14.00	79.07	13.65	.003	.098	.31**
		T2	64.89	13.57	74.73	13.37	.001	.120	.35**
	Utility	T1	56.26	27.93	69.69	22.87	.016	.067	.26*
		T2	61.40	18.90	65.17	17.07	.334	.011	.10
Learning	Knowledge	T1	67.25	12.25	73.58	14.36	.031	.054	.23*
		T2	65.17	13.24	70.73	12.95	.053	.044	.21
	Self-efficacy	T1	40.72	21.02	55.43	24.50	.004	.095	.31**
		T2	46.75	17.59	54.57	17.89	.044	.047	.22*
Behavior	Application to practice	T1	47.89	24.89	59.99	25.28	.028	.056	.24*
		T2	53.36	15.91	58.53	15.53	.131	.027	.16
Results	Organizational results	T1	47.78	20.62	60.32	20.09	.005	.088	.30**
		T2	48.25	16.94	57.71	16.25	.010	.077	.28**

Note. For objective training success, means of the performance test for factual knowledge ranged from 0 to 57 points and means of the performance test for applied knowledge ranged from 0 to 20 points. For subjective training success, means of the Q4TE on Likert scale ranged from 0% (*completely disagree*) to 100% (*completely agree*). *M* = mean; *SD* = standard deviation; η^2 = effect size with $.01 \le \eta^2 < .06$ (small), $.06 \le \eta^2 < .14$ (moderate), $\eta^2 \ge .14$ (large) according to Cohen (1988); *r* = correlation between training success measures (performance test or Q4TE scales) and training setting group with 1 = e-learning and 2 = classroom training. * Correlation is significant at the .05 level (two-tailed). ** Correlation is significant at the .01 level (two-tailed).

Table 29. *Correlations for Objective and Subjective Training Success Measures Immediately After Training in Time 1*

	1	2	3	4	5	6	7	8
1 Factual knowledge	-	.57**	-.24	.20	-.24	-.05	-.05	.02
2 Applied knowledge	.53**	-	.04	.21	.07	.05	.14	.21
3 Satisfaction	.16	-.12	-	.43**	.79**	.46**	.40**	.43**
4 Utility	.21	-.02	.35*	-	.40**	.62**	.52**	.63**
5 Knowledge	.18	-.18	.75**	.57**	-	.59**	.51**	.57**
6 Self-efficacy	.21	.02	.21	.53**	.59**	-	.64**	.87**
7 Application to practice	.26	.05	.26	.66**	.59**	.81**	-	.76**
8 Organizational results	.21	.00	.35*	.67**	.63**	.88**	.84**	-

Note. For e-learning, correlations are shown above the diagonal and for classroom training, correlations are shown below the diagonal. * Correlation is significant at the .05 level (two-tailed). ** Correlation is significant at the .01 level (two-tailed).

Table 30. *Correlations for Objective and Subjective Training Success Measures Six to Eight Weeks Later in Time 2*

	1	2	3	4	5	6	7	8
1 Factual knowledge	-	.49**	-.08	.08	-.06	-.03	.09	-.18
2 Applied knowledge	.48**	-	.20	.22	.33*	.18	.30	.21
3 Satisfaction	.33*	.21	-	.68**	.86**	.33*	.43**	.32*
4 Utility	.10	.09	.63**	-	.72**	.43**	.50**	.38*
5 Knowledge	.24	.18	.87**	.76**	-	.52**	.62**	.53**
6 Self-efficacy	.16	.12	.28	.46**	.44**	-	.76**	.84**
7 Application to practice	.20	.17	.41**	.62**	.59**	.82**	-	.81**
8 Organizational results	.17	.17	.33*	.52**	.52**	.93**	.84**	-

Note. For e-learning, correlations are shown above the diagonal and for classroom training, correlations are shown below the diagonal. * Correlation is significant at the .05 level (two-tailed). ** Correlation is significant at the .01 level (two-tailed).

Subjective training success

Comparison of training settings. Training success was assessed subjectively in terms of trainees' self-reporting on various training success scales using the Q4TE separately for both training settings. At an introductory level, descriptive statistics showed similarities between both training settings. Regarding both time periods, means of the six Q4TE scales that cover also all levels of Kirkpatrick's (1959) framework showed the same sequence in both training settings: *Satisfaction* always scored highest, followed by *knowledge*, and *utility*. *Application to practice*, *organizational results*, and *self-efficacy* in this sequence always scored lowest independently of training setting.

A mixed between-within subjects multivariate analysis of variance was performed to assess the impact of the two different training settings (e-learning, classroom training) on trainees' scores on the Q4TE across two time periods (immediately after training and six to eight weeks follow-up). All six Q4TE scales served as dependent variables. This analysis investigated the question whether changes in subjective training success scores over time were identical for the two training setting groups. There was no significant interaction between training setting and time, Wilks' Lambda = .93, $F(6, 79) = .94$, $p = .47$, $\eta^2_{partial} = .067$. There was a substantial (very large) main effect for training setting, Wilks' Lambda = .73, $F(6, 79) = 4.98$, $p < .001$, $\eta^2_{partial} = .28$, indicating that there was a significant difference between e-learning and classroom training on the combined dependent training success variables. When the results for the dependent variables were considered separately, all Q4TE scales showed moderate to large training setting effects with an alpha level of .05 (*satisfaction*: $p < .001$, $\eta^2_{partial} = .15$, large training setting effect; *utility*: $p = .017$, $\eta^2_{partial} = .066$, moderate training setting effect; *knowledge*: $p = .019$, $\eta^2_{partial} = .063$, moderate training setting effect; *self-efficacy*: $p = .002$, $\eta^2_{partial} = .11$, moderate training setting effect; *application to practice*: $p = .018$, $\eta^2_{partial} = .065$, moderate training setting effect; *organizational results*: $p < .001$, $\eta^2_{partial} = .12$, moderate training setting effect). An inspection of the mean scores indicated that e-learning trainees reported lower scores on all training success scales compared to classroom training trainees. There was also a large main effect for time, Wilks' Lambda = .81, $F(6, 79) = 3.19$, $p = .007$, $\eta^2_{partial} = .20$, indicating that there was a significant difference between the two time periods on the combined dependent training success variables. When the results for the dependent variables were considered separately with an alpha level of .05, scores significantly differed only on *satisfaction* ($p = .005$, $\eta^2_{partial} = .091$, moderate time effect). All other time effects were very small to small (all ps ≥ .079, all η^2s ≤ .036). Inspecting the mean scores combined across both training settings, some training success scales showed a reduction, some showed an increase from Time 1 to Time 2.

To further identify specific differences between e-learning and classroom training depending on the time period, follow-up univariate analyses were conducted (see Figure 24). Regarding all training success scales, trainees participating in classroom training always scored higher compared to e-learning trainees immediately after training in Time 1 with moderate training setting effects (all ps ≤ .031, all η^2s ≥ .054). The greatest differences occurred on the scales: *satisfaction* ($p = .003$, $\eta^2 = .098$, moderate training setting effect), *self-efficacy* (p = .004, η^2 = .095, moderate training setting effect), and *organizational results* (p = .005, η^2 = .088, moderate training setting effect).

Focusing on the short-term evaluation outcome *knowledge* (measuring success of learning) and focusing on the long-term evaluation outcome *application to practice* (measuring success of transfer) immediately after training in Time 1, trainees participating in classroom training scored higher compared to e-learning trainees with moderate training setting effects (*knowledge*: p < .031, η^2 = .054; *application to practice*: p = .028, η^2 = .056). Thus, Hypothesis 3a and Hypothesis 4a were supported.

However, six to eight weeks later in Time 2, these moderate training setting effects were reduced indicating that the differences between e-learning and classroom training were reduced across time (for detailed differences between training settings separately for each time period and Q4TE scale, see Table 28). Regarding success of learning (*knowledge*) and success of transfer (*application to practice*) six to eight weeks later in Time 2, there were no significant differences between training settings. Therefore, Hypothesis 3b and Hypothesis 4b were rejected (*knowledge*: p = .053, η^2 = .044, small training setting effect; *application to practice*: p = .13, η^2 = .027, small training setting effect).

Comparison of time periods. To further investigate the overall time effect of the mixed between-within subjects multivariate analysis of variance, follow-up univariate analyses/paired-samples *t*-tests were conducted. In e-learning, the means of the Q4TE scales descriptively increased across time (besides *satisfaction* and *knowledge*), whereas in classroom training, all Q4TE means descriptively decreased across time. Only scores on *satisfaction* showed a significant decrease across time in both training settings (e-learning: p = .043, η^2 = .034, small time effect; classroom: p = .052, η^2 = .025, small time effect). All other time effects were very small to small in both training settings (all ps ≥ .15, all η^2s ≤ .024). These results were in line with the previous mixed between-within subjects multivariate analysis of variance. Success of learning (e-learning: p = .30, η^2 = .007, no time effect; classroom: p = .15, η^2 = .011, small time effect) and success of transfer (e-learning: p = .19, η^2 = .017, small time effect; classroom: p = .70, η^2 = .001, no time effect) perceived by trainees did not statistically reduce across time in both training settings; thus Hypothesis 5a and Hypothesis 5b were not supported.

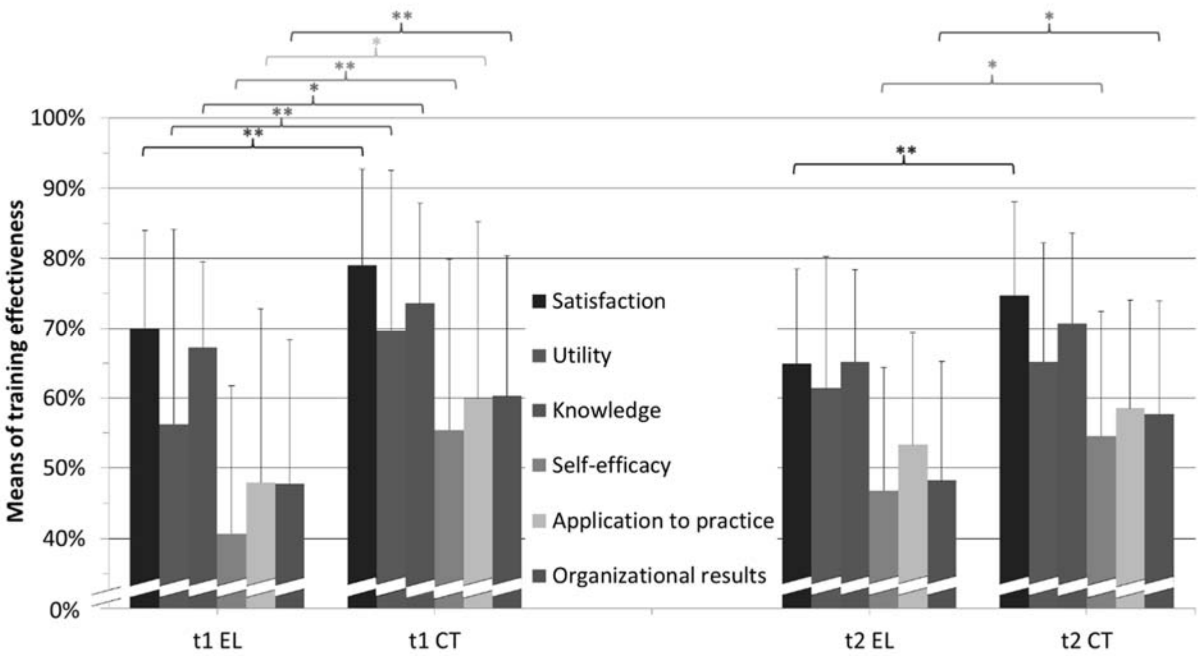

Figure 24. Means of Q4TE scales measuring training effectiveness comparing training settings immediately after training in Time 1 and six to eight weeks later in Time 2 with EL = e-learning and CT = classroom training. Mean on Likert scale ranging from 0% (*completely disagree*) to 100% (*completely agree*). * $p < .05$ with $.01 \leq \eta^2 < .06$ (small). ** $p < .01$ with $.06 \leq \eta^2 < .14$ (moderate) according to Cohen (1988).

5.2.4 Discussion

Summary

The fifth study aimed at contributing to filling the specific research gap regarding the comparison of training success between corporate e-learning and classroom training. While training contents were equal (anatomy and pathology of anesthesia and emergency medicine) and requirements of a good transfer-supportive training setting were fulfilled as checked by T&D professionals, only training settings differed: E-learning was rather flexible as to time and place and not guided by a trainer, whereas traditional classroom training was fixed in time and in location, clearly structured, and guided by a trainer.

This field experiment (i.e., trainees were randomly assigned to the two training settings) and time-lag study systematically examined differences between e-learning and classroom training regarding objective as well as subjective training success. To obtain clear training setting effects, we controlled for Big Five personality traits, proactive personality, transfer factors (indicating an equal transfer climate), gender, and age. As these variables were stable across time and did not differ across training settings, effects result from the specific training setting.

First, as to objective training success, both training settings did not differ in factual knowledge right after the training course. However, patterns of factual performance scores developed differently in both training settings over time, indicating small time effects: Whereas e-learning trainees significantly improved, the scores of classroom trainees significantly decreased. Six to eight weeks after the training course, factual knowledge had a greater effect in e-learning compared to classroom training settings, confirming Hypothesis 1b.

Analyzing applied knowledge, the reversed pattern for each time period appeared: Whereas training settings differed right after the training course (large training setting effect in favor of classroom), six to eight weeks later, applied knowledge had a similar effect in e-learning as well as in classroom training, confirming Hypothesis 2b. Again, e-learning trainees significantly improved and classroom training trainees significantly decreased in applied performance scores with moderate time effects. In sum, our results for factual and applied knowledge are in line with results found by the MMB (2010) and Sitzmann et al. (2006).

Second, analyzing subjective training success descriptively in both time periods, both training settings showed the same sequence in means of the Q4TE scales covering all levels of Kirkpatrick's (1959) framework: *Satisfaction* always scored highest, followed by *knowledge*, *utility*, *application to practice*, *organizational results*, and *self-efficacy*. These results are in line with Alliger et al. (1997) and Tharenou et al. (2007), demonstrating that affective reactions (*satisfaction*) often received relatively high values which might be due to trainees' feeling of appreciation to participate in training, but they did not (necessarily) result in high values of success of transfer. Further, as also described in the research literature, success of transfer is lower than success of learning (all ps < .001). Generally, our study replicated means of the Q4TE from comparative studies investigating 600 trainees from diverse companies (Kauffeld, 2010).

Analyzing subjective training success inductively, training success was only perceived differently right after the training course. Later on, it was perceived equally effective in both training settings. In support of Hypothesis 3a and Hypothesis 4a, regarding all training success scales (including our focused evaluation outcomes success of learning and success of transfer) trainees participating in classroom training scored higher compared to e-learning trainees with moderate training setting effects right after the training course. However, six to eight weeks later, these training setting effects disappeared as success of learning and success of transfer were similarly perceived by both groups of trainees. Hypothesis 3b and Hypothesis 4b had to be rejected. Thus, the perception of training success changed across time.

Further, investigating time effects for each training setting separately, results showed very small to small time effects for all Q4TE scales in both training settings (all ps ≥ .15, all η^2s ≤ .024). Interestingly, in e-learning, Q4TE means descriptively increased across time (besides *satisfaction*), whereas in classroom training, all Q4TE means descriptively decreased across time which indicated that the transfer problem (e.g., Baldwin & Ford, 1988; Grossman & Salas, 2011; Saks et al., 2014) already occurred within six to eight weeks after training (in classroom training only). As success of learning and success of transfer were not statistically reduced across time, the spotlight of our evaluation outcomes remained stable, and Hypothesis 5a and Hypothesis 5b were not supported.

To interpret the effects found in our study (i.e., differences across training settings, increase in e-learning, and decrease in classroom training), various explanations are conceivable: First, the reason for higher scores in factual knowledge in e-learning trainees compared to classroom trainees might be explained by the phenomenon of environmental context-dependent memory (Smith, 1994; Smith & Vela, 2001). That is, taking a test in the same external learning environment will make it easier to retrieve those memories. For e-learning trainees, the training and retrieval context were both online (congruent). Thus, more autonomous application was possible here, whereas for classroom trainees, the training context was face-to-face and the retrieval context was online (incongruent). It has been demonstrated that environmental context-dependent memory works well especially for factual knowledge (e.g., Godden & Baddeley, 1975). However, for high scores in applied knowledge, further driving factors are necessary.

Second, as e-learning trainees did not receive any feedback from the trainer or other trainees and no trainer guided trainees during training (e.g., by clarifying any questions), e-learning trainees experienced a more difficult application in the performance test. Thus, they might have obtained lower scores in applied knowledge compared to classroom trainees right after training.

Third, objective and subjective training success measures interact via self-concept. This issues refers directly to the bidirectional relationship between self-concept and performance (Kurtz-Costes & Schneider, 1994; Hansford & Hattie, 1982). For example, as implicit and explicit feedback received in the performance test can trigger a certain self-concept which in turn can serve as a precursor for subjective training success measures, trainees assessed themselves according to their perceived and feedbacked performance in the test. Consequently, e-learning trainees obtained lower scores in subjective training success measures compared to classroom trainees right after training.

Fourth, it might be the case that e-learning trainees rated themselves lower in subjective training success immediately after training exposure because they had less confidence in the source of information (e-learning setting). This effect disappeared as people rather forget the source of information than the information itself, thus making the source irrelevant. This result refers to the sleeper effect; that is, "when people receive a communication associated with a discounting cue, such as a noncredible source, they are less persuaded immediately after exposure than they are later in time" (Kumkale & Albarracín, 2004, p. 143).

Fifth, since e-learning trainees might have had the feeling that they did not achieve high levels of success of learning and success of transfer after training and/or the first assessment, they showed a different behavior (additionally assessed in the questionnaire) compared to classroom trainees during assessment one and two: E-learning trainees (a) more often talked about the training contents with colleagues, (b) more often repeated the training contents, (c) more often applied the context learned in training, and (d) experienced retrospectively that the training was useful for their work. Supposedly, trainees behaved differently to compensate for their feelings of unease.

To further explain the increase in e-learning and decrease in classroom training, again, environmental context-dependent memory applies (Smith, 1994; Smith & Vela, 2001). Consequently, e-learning trainees showed an improved recall of information which results in a gain in objective training success (factual as well as applied knowledge). Then again, this perceived gain in performance scores can trigger a higher self-concept including a gain in confidence which can result in a gain in subjective training success in e-learning trainees. As a result of behavior during assessments, training success of e-learning trainees can be described as a gain across time, thus catching up on previous large differences in applied knowledge and moderate differences in subjective training success right after the training course.

Limitations

Content limitations are discussed first followed by methodological limitations. To further investigate what happened between both assessments, gathering data about further control variables in subsequent studies seems beneficial, such as variable measuring behavior during assessments (e.g., how often did you talk about the training content with supervisors). Also, pre-knowledge can serve as a further control variable as prior knowledge is crucial to results in performance tests (for further literature on expertise and pre-knowledge, see Schneider, 2000, 2008). For example, assessing pre-knowledge at a baseline level seems important because differences in performance can be more clearly attributed to the training.

Methodological limitations refer to the survey format and selected sample. Regarding objective training success, as standard deviations did not vary to a great deal, and data was not normally distributed (all *p*s $\leq .05$ suggesting violation of normality), a performance test with items varying in levels of difficulty seems more convenient. Further, a rather balanced award for points for factual and applied knowledge helps to better compare both types of knowledge. On purpose, we did not calculate a performance sum score since the focus of the investigation was on differences between factual and applied knowledge. In addition, objective training success measures including *organizational results* (e.g., turnover) are preferable, but they often depend on other extraneous environmental factors that are not related to the training itself (Goldstein & Ford, 2002).

Subjective training success was assessed with self-reporting of trainees using the widely implemented questionnaire Q4TE (Grohmann & Kauffeld, 2013; Kauffeld et al., 2009). Thus, being aware of the common method bias, that is, variance that is attributable to the measurement method rather than to the constructs the measures represent (e.g., Podsakoff, MacKenzie, Lee, & Podsakoff, 2003; Spector, 2006), gathering data with multiple methods would be an asset. Multi-dimensional assessments should be investigated, for example, by involving not only trainees but also supervisors, colleagues of trainees, or customers.

This study was designed as a time-lag study with a time period of six to eight weeks after the training courses. A longitudinal study assessing objective and subjective training success on a longer run with more points of time would be desirable in order (a) to facilitate a more frequent application of transfer activities, (b) to see how training success develops across time, or (c) to check whether organizational goals were achieved.

Methodologically, Bonferroni alpha adjustment is necessary when interpreting results. However, as we conducted many *t*-tests, this alpha level would have been reduced to .004, and therefore, it is very conservative in its nature. Since our sample size in each group was restricted due to trainees available at the company, resulting in a relatively small sample, Bonferroni adjustment seems too conservative.

In summary, to confirm our results and to gain a higher impact, more trainees employed at companies which are in diverse branches of industry are needed in future studies to enlarge the sample size, so that implications have a wider scope of application. Practical implications are discussed and further research objectives are suggested.

Implications for practice

The study reacted to the claim that effectiveness studies examining predictors of success of learning and success of transfer as facets of training success are needed (Aguinis & Kraiger, 2009). Investigating an overall performance sum score, none of both training settings dominates the other, but focusing on specific training contents, we confirmed that declarative knowledge is better learned in e-learning, whereas procedural knowledge is better learned in classroom training right after a training course. Regarding the connectedness between factual and applied knowledge, factual knowledge is rather a precursor for applied knowledge, and applied knowledge is the key outcome in the work context. Therefore, trainees need to have a proper factual knowledge basis to actually be successful in transfer of learning. Substantial strengths of classroom training are: (a) more realistic exercises due to face-to-face contact, (b) easier recognition and imitation of emotions, (c) discussion of errors that will help learn from one's mistakes or from mistakes of others, and (d) an increased possibility to ask questions and to get immediate answers from the trainer. Systematically combining strengths of e-learning and classroom training, blended-learning seems very promising (Kaltenbaek, 2003). For example, when teaching factual knowledge, use e-learning combined with follow-up classroom training to apply what you have learned practically. Another option for previous e-learning training (e.g., teaching factual knowledge) is providing subsequent feedback from trainer and/or colleagues.

For T&D professionals constructing the e-learning program, it is important to sufficiently design the learning platform by considering transfer-supportive measures (Salas et al.,2012), then, e-learning can serve as an adequate training setting. Further, especially e-learning trainees obtained low levels in self-efficacy (even though there was an increase across time) which can be due to less feedback during training from the trainer and/or other trainees. Trainees with a better knowledge and a more frequent use of self-regulated learning strategies were better in decoding and recoding knowledge and skills learned in e-learning (software training) and they also showed higher levels of self-efficacy (Gravill & Compeau, 2008). Thus, it seems important for T&D professionals to support their trainees in developing and engaging in self-regulated learning strategies to better coordinate the individual learning process and to achieve learning objectives adequately.

In sum, strengths and weaknesses of both training setting groups balance out six to eight weeks after training. Thus, none of the training settings dominates the other. Consequently, it is not the specific training setting that is particularly promising in the end. But which other factors substantially enhance (or inhibit) training success? The influence of factors affecting training transfer is specifically investigated in the next study.

5.3 Exploring transfer factors to ensure training success in corporate e-learning

5.3.1 Aims and research questions

Aims

Increasing evidence shows that more and more companies are investing in extensive HRD activities to survive the “war for talent”. For instance, the State of the Industry Report by the Association for Talent Development found that U.S. organizations spent $125.8 billion in 2009 and $156.2 billion on employee learning and development in 2011 (ASTD, 2012, para. 2). Similarly, €28.6 billion was spent on training of employees in 2010 in Germany (Seyda & Werner, 2012) which is a nominal growth of 6.4% compared to 2007 (Lenske & Werner, 2009).

However, according to Holton (2015, “United States Training Industry”), only 10-30% result in changed job performance. Indeed, companies often observe that the knowledge and skills acquired in training are insufficiently transferred to the workplace (Grossman & Salas, 2011). “If we cannot deliver transfer, then training is wasted” (Holton, 2015, “The Future of Training”). Training transfer is defined as “the degree to which trainees effectively apply the knowledge, skills, and attitudes gained in a training context to the job” (Baldwin & Ford, 1988, p. 63). For example, if the training situation differs greatly from the work situation, a phenomenon called *transfer problem* very likely occurs (Baldwin & Ford, 1988; Saks et al., 2014). This effect is confirmed by several scientific studies including meta-analyses (Grossman & Salas, 2011; Saks et al., 2014). Training transfer is a current and important topic: The International Journal of Training and Development recently published its first special issue on training transfer (Saks et al., 2014). Many theoretical models, for example, the HRD evaluation and research model by Holton (2005) or the transfer process model by Baldwin and Ford (1988), and many empirical studies (mainly focusing on classroom training) investigated transfer-related factors in the person, training, and organization that affect training transfer (Holton, 2005). Hence, it is crucial to understand how training transfers into job performance.

Eversince technological developments have enabled e-learning, an increasing number of companies has shifted from traditional classroom training to e-learning, as it is less costly. E-learning is defined as "the use of electronic technologies to deliver information and facilitate the development of skills and knowledge" (ASTD, 2012, “Content distribution,” para. 6). Technology-based training settings account for 37.3% of hours spent on formal training (ASTD, 2012, “Content distribution,” para. 6). Also in Germany, about every fourth company implemented computer-based learning referring to computer-based training, web-based training, and e-learning (Seyda & Werner, 2012). In fact, more than two-thirds (72%) of the companies that have been awarded as “best employers in Germany” offered e-learning as part of their HRD activities (for details, see section 5.1). To date, only few studies explored corporate e-learning settings with its transfer factors. The aim of this study is to shed light on the transfer problem in corporate e-learning settings, thus combining both issues.

As to important training outcome measures according to Kirkpatrick’s level framework (1959), research demonstrated that *reactions* (Level 1) and *organizational results* (Level 4) are not considered as sufficient because (a) trainees’ satisfaction with the training does not necessarily correlate with success of learning or success of transfer (Alliger et al., 1997), and (b) it is quite difficult for trainees to gather valid and realistic data about organizational results (Goldstein & Ford, 2002).

Transfer is the ultimate outcome in the training process and serves as the dependent variable in most studies (e.g., Baldwin & Ford, 1988; Ford & Weissbein, 1997; Grossman & Salas, 2011). According to the four-level evaluation model (Kirkpartick, 1959) transfer refers to Level 3: behavior which measures the extent and quality to what trainees apply the newly acquired knowledge, skills, and attitudes to their daily work. As the model describes the levels as hierarchically and causally sequenced, learning (Level 2) is understood as a precursor of transfer which is also in line with Baldwin and Ford’s (1988) transfer process model. Further, with reference to the revised HRD evaluation and research model by Holton (2005), learning precedes individual performance that refers to training transfer. Hence, it seems beneficial to simultaneously consider success of learning and success of transfer as important outcome variables. As transfer factors in Holton’s (2005) model show direct and indirect effects via learning on transfer, these effects are further inspected in this study.

A large variety of factors affecting training transfer have been identified in a vast amount of research studies and theoretical models (e.g., Baldwin & Ford, 1988; Ford & Weissbein, 1997; Kraiger et al., 1993; Holton, 1996; Noe, 1986; Noe & Schmitt, 1986; Steiner, Dobbins, &Trahan, 1991). Importantly, a psychometrically sound measure is a prerequisite to validly study factors affecting training transfer. Grounded in training transfer research and based on Holton’s revised HRD evaluation and research model (2005), the LTSI (Holton et al.,

2000) seems to be the leading diagnostic instrument assessing factors that influence training transfer (Chen et al., 2005).

Previous research studies on important transfer factors especially in e-learning settings found that environmental-related factors (e.g., supervisor and co-worker support) are the largest contributors to the prediction of training transfer (Bates et al., 2000). Another study been demonstrated that work environment factors and individual attitudes explained most of the variance in motivation to transfer training in a computer-based training (Seyler, Holton, Bates, Burnett, & Carvalho, 1998). In a comprehensive review on the concept of motivation to transfer, Gegenfurtner et al. (2009) described the central role of motivation to transfer in the transfer process. Here, motivation to transfer was identified as a mediator of antecedents such as trainee characteristics on training transfer which was replicated in a recent study (Grohmann et al., 2014).

Research Questions

To shed light on factors affecting training transfer and to contribute to the transfer problem literature, this study investigates which transfer factors are substantial for training success in corporate e-learning settings. Training success is explored in terms of success of learning and success of transfer. In addition to e-learning settings, corporate classroom training were investigated to compare results found for corporate e-learning settings and to identify possible differences in patterns. This leads to the following research questions:

Research question 1: Which transfer factors are important for success of learning in corporate e-learning (1a) and classroom training settings (1b)?

Research question 2: Which transfer factors are important for success of transfer in corporate e-learning (2a) and classroom training settings (2b)?

Combining the results from both research questions, the mediating role of success of learning in the relationship between transfer factors and success of transfer was explored in a path model separately for corporate e-learning and classroom training settings.

Research question 3: Which transfer factors indirectly affect success of transfer via success of learning and which transfer factors affect it directly in corporate e-learning (3a) and classroom training settings (3b)?

For reasons of simplification, learning refers to success of learning, and transfer refers to success of transfer in the following sections on methodology and results.

5.3.2 Methodology

Participants

The sample was drawn from several companies and trainings to get an insightful view on e-learning. The overall sample stems from six different German large-scale enterprises (IfM-Institut für Mittelstandsforschung Bonn, 2014) which varied in branches of industry. Investigating the employees' view on training transfer, the overall sample included k = 12 distinct samples and comprised a total number of N = 974 trainees (see Table 31). In this study, the meta-analytic approach was applied to aggregate information on factors affecting training transfer and to achieve higher statistical power compared to measures derived from a single study. The study aimed at combining results from different studies to identify patterns among study results. Thus, the meta-analytic approach was applied as an appropriate way to summarize information from our several single studies. All studies were conducted in a corporate environment. The majority of studies (all except one) were cross-sectional surveys. The training contents and training duration differed. Besides primarily investigating established e-learning settings, also new forms of e-learning, that is, virtual classrooms and online communities, were integrated in this meta-analysis. In all studies, trainees rated either the sum of all e-learning trainings (true for the majority of samples) or they rated a specific e-learning or classroom training they had recently participated in. Weighted by samples size, trainees were on average 36.15 years old (SD = 7.46) and the samples consisted on average of 53% men (SD = 9%).

Table 31. *Overview of the Studies Included in the Meta-Analysis*

Sample ID	Study sample			
(Study ID)	*N*	Branch of industry	Training setting	Males
1 (1)	329	Engineering	E-learning	49%
2 (2)	119	Manufacturing	E-learning	-
3 (2)	75	Manufacturing	E-learning	41%
4 (4)	68	Automotive	Online communities	49%
5 (2)	51	Manufacturing	Blended learning	55%
6 (3)	43	Engineering	E-learning	67%
7 (5)	41	Manufacturing	E-learning	66%
8 (6)	39	Manufacturing	Virtual classroom	85%
9 (4)	33	Automotive	Virtual classroom	49%
10 (4)	104	Automotive	Classroom training	49%
11 (5)	45	Manufacturing	Classroom training	56%
12 (2)	27	Manufacturing	Classroom training	59%

Note. N = number of sample sizes.

Instruments

Training success. The Questionnaire for Professional Training Evaluation (initial Q4TE: Kauffeld et al., 2009; Q4TE: Grohmann & Kauffeld, 2013) was used to assess training success. It has been demonstrated that the Q4TE is time-efficient, psychometrically sound, and widely applicable across different training contents and training settings (Grohmann & Kauffeld, 2013). The subscale *knowledge* assesses learning, whereas the subscale *application to practice* is equivalent to transfer. Based on self-report measures, trainees rated 7 items on an 11-point response scale ranging from 0% (*completely disagree*) to 100% (*completely agree*) with single steps of 10% increase each (for sample items and reliabilities, see Table 27 in section 5.2.2).

Transfer factors. The German version of the LTSI (LTSI; Holton et al., 2000; GLTSI; Kauffeld et al., 2008) was applied to assess transfer factors. The LTSI is based on the HRD evaluation and research model (Holton, 1996, 2005) and measures factors that influence training transfer. The LTSI is a comprehensive, standardized psychometric inventory and received strong empirical evidence of construct validity (e.g., Holton et al., 2000; Holton et al., 2007), criterion validity (e.g., Ruona et al., 2002), and cross-cultural validity (e.g., Chen et al., 2005 [Taiwan version]; Devos et al., 2007 [French version]; Kauffeld et al., 2008 [German version]; Khasawneh et al., 2006 [Arabic version]; Yamkovenko et al., 2007 [Ukrainian version]). There is evidence of the questionnaire's generalizability across training programs as heterogeneous samples were assessed, ranging from government, public for-profit, private, and non-profit organizations (Holton et al., 2000). A recent global validation study (N = 6,120 participants) conducted in 17 countries (very diverse) and 14 different languages showed that the LTSI scales work across cultures and languages (Holton, 2015, "Latest Developments"). The GLTSI was also validated psychometrically (Bates et al., 2007; Kauffeld et al., 2008). In total, the GLTSI comprises 67 items measuring 16 constructs that are identified as critical barriers and catalysts of transfer (Wirth et al., 2009). As presented in Table 32, the GLTSI factors are categorized in the domains ability, motivation, and work environment, and they either refer to a specific training program (11 factors with 44 items) or to training in general (5 scales with 23 items) according to Kauffeld et al. (2008). Participants rated the items on a 5-point Likert scale (1 = *strongly disagree*, 2 = *disagree*, 3 = *neither agree nor disagree*, 4 = *agree*, and 5 = *strongly agree*).

Table 32. *Scales and Sample Items of the LTSI (Holton et al., 2000, pp. 344-346)*

Scale	s/g	α	*n*	Sample item
Ability				
Perceived content validity	s	.86	5	What is taught in training closely matches my job requirements.
Transfer design	s	.82	4	The activities and exercises the trainers used helped me know how to apply my learning on the job.
Personal capacity for transfer	s	.65	4	My workload allows me time to try the new things I have learned.
(Opportunity to use)	s	.58	4	The resources I need to use what I learned will be available to me after training.
Motivation				
Motivation to transfer	s	.88	4	I get excited when I think about trying to use my new learning on my job.
Learner readiness*	s	.82	4	Before the training I had a good understanding of how it would fit my job-related development.
Transfer effort–performance expectations	g	.78	4	My job performance improves when I use new things that I have learned.
Performance–outcomes expectations	g	.87	5	When I do things to improve my performance, good things happen to me.
Performance self-efficacy*	g	.79	4	I am confident in my ability to use newly learned skills on the job.
Work environment				
Peer support	s	.82	4	My colleagues encourage me to use the skills I have learned in training.
Supervisor support	s	.89	5	My supervisor sets goals for me that encourage me to apply my training on the job.
(Supervisor sanctions)	s	.71	3	My supervisor opposes the use of the techniques I learned in training.
(Positive personal outcomes)	s	.78	3	Employees in this organization receive various “perks” when they utilize newly learned skills on the job.
(Negative personal outcomes)	s	.85	4	If I do not utilize my training I will be cautioned about it.
Openness to change	g	.73	5	People in my group are open to changing the way they do things.
Performance coaching	g	.80	4	After training, I get feedback from people about how well I am applying what I learned.

Note. Excluded GLTSI factors are put in brackets. s = training-specific scales (11); g = general scales (5). α = Cronbach’s α of GLTSI scales (Kauffeld et al., 2008). *n* = number of items of the total GLTSI. * Scales refer to secondary influences.

Procedures

Coding. For each sample, sample characteristics were coded such as branch of industry, sample size, and the specific training setting (see Table 31). Pearson's correlation coefficient *r* between each transfer factor (GLTSI factors) and each training success measure (learning and transfer) was coded for the meta-analysis. Finally, 12 GLTSI factors with learning and transfer were coded (excluded GLTSI factors: *opportunity to use*, *supervisor sanctions*, *positive personal outcomes*, *negative personal outcomes*).

Data Analysis Techniques. Multiple studies aggregate information on focused measures to achieve higher statistical power compared to measures derived from a single study. Meta-analyses combine results from different studies to identify patterns among study results. Hence, they are an appropriate way to summarize information from our several single studies.

Further, meta-analytic structural equation modeling (MASEM) was used (a) to assess what actually drives learning and transfer, when predictors are overlapping and (b) to build a more complex mediation model. According to Geiser (2010), MASEM combines classical meta-analysis with structural equation modeling as it allows for direct comparisons of the strength of two parallel effects or the testing of complex models with several interdependent outcomes. The two-step MASEM procedure proposed by Cheung and Chan (2005) was conducted:

First, sample-size-weighted effect sizes (ES *r*) using random effects models were calculated and inspected for significance and homogeneity. For e-learning and classroom training settings, a significance level of $p \leq .05$ and a homogeneity level of $p > .05$ were applied. If both inclusion criteria were fulfilled, these overall effect sizes served as input information for further analyses in MASEM.

Second, the resulting correlation matrix of all variables (GLTSI factors with learning, GLTSI factors with transfer, and GLTSI factors among each other) and the harmonic mean calculated from the overall sample size associated with each meta-analytic correlation were used in structural equation modeling as suggested by Landis (2013).

Data Analysis. Within the described MASEM framework, our research questions, namely, (a) which transfer factors are important for learning and (b) which transfer factors are important for transfer, were tested first by inspecting the significance and homogeneity of the sample-size-weighted effect sizes (meta-analytic correlations), separately for corporate e-learning and classroom training settings. Analyses revealed that (a) many GLTSI factors were significant and homogeneous and (b) GLTSI factors are not independent. Consequently, overlapping effects are very likely to occur if GLTSI factors on learning and on transfer are included in one (too complex) overall structural equation model. Being aware of overlapping effects, therefore, the significant and homogeneous subset of transfer factors was entered into saturated path models investigating direct effects of GLTSI factors (independent variables) on learning (dependent variable) and separately on transfer (dependent variable). Hence, key factors predicting training success are extracted in these analyses, namely those, whose path coefficients are significant. Finally, we are able to satisfactorily answer the first two research questions and to identify what drives learning and transfer, respectively.

Second, to investigate the complex relationship between transfer factors, learning, and transfer, a mediation model including all drivers identified in the first step was conducted separately for corporate e-learning and classroom training settings. Previous analyses showed that learning was strongly related to transfer (e-learning: ES r = .67, p < .001; classroom: ES r = .63, p < .001). The mediation model tested for those transfer factors which were significant for either learning or transfer in the saturated path models. These analyses help answer the third research question which transfer factors affect transfer directly and/or indirectly.

Third, to compared e-learning and classroom training settings, the strength of the meta-analytic correlations from these independent samples was examined in contrast across training settings according to Eid, Gollwitzer, and Schmitt (2011, pp. 547) using the psychometrica website (Lenhard & Lenhard, 2014).

5.3.3 Results

Meta-analytic correlations

As indicators of overall effect sizes, meta-analytic correlations regarding the relationship between transfer factors and training success are presented, separately for learning and transfer as an outcome.

E-learning and learning. Calculating the combined, meta-analytic correlations between the 12 transfer factors and learning, seven associations were significant (all *p*s < .001) and homogeneous (all *p*s ≥ .069 for Q statistics; for detailed statistical coefficients, see Table 33). *Motivation to transfer* and *transfer design* showed large-sized effects (Cohen, 1992) on learning. *Transfer effort–performance expectations*, *perceived content validity*, *learner readiness*, and *peer support* had medium-sized effects, whereas *performance self-efficacy* showed a small effect on learning. Furthermore, all corresponding 95% confidence intervals (CI) excluded zero, thus attesting to the statistical significance of the meta-analytic correlations. As all of these seven transfer factors were identified as possible predictors of learning, they were included in a structural path model simultaneously to establish drivers of learning. Results of the MASEM (Model 1, harmonic mean = 504) demonstrated that *motivation to transfer* showed the strongest influence on learning (r = .46, p < .001). *Transfer design* was identified as the second strongest influence on learning (r = .28, p < .001), all other predictors exhibited non-significant paths (all other *p*s ≥ .11).

E-learning and transfer. Investigating transfer as an outcome in e-learning settings, five of the 12 associations between transfer factors and transfer met both inclusion criteria of significance (all *p*s ≤ .010) and homogeneity (all *p*s ≥ .075 for Q statistics; for detailed statistical coefficients, see Table 33). Generally, meta-analytic correlations were descriptively larger for transfer compared to learning. In accordance to learning as outcome in e-learning, the same significant transfer factors were identified and also *motivation to transfer* and *transfer design* revealed large effects on transfer, *perceived content validity* and *learner readiness* showed medium effects, whereas *performance self-efficacy* exhibited only a small effect on transfer. Results of the MASEM (Model 2, harmonic mean = 506) revealed that, again, *motivation to transfer* showed the strongest influence on transfer (r = .39, p < .001), and *transfer design* showed the second strongest influence on transfer (r = .24, p < .001). Additionally, *learner readiness* was identified as relevant to transfer (r = .092, p = .024), whereas all other predictors had non-significant paths to transfer (all other ps ≥ .11).

Classroom training and learning. Meta-analytic correlations between the 12 transfer factors and transfer showed that five associations reached statistical significance (all ps ≤ .036) and were homogeneous (all ps ≥ .31 for Q statistics; for detailed statistical coefficients, see Table 33). Results showed a large effect for *transfer design* and a medium effect for *transfer effort–performance expectations*, whereas small effects for *motivation to transfer*, *performance coaching*, and *performance self-efficacy* on learning were found. Results of the MASEM (Model 3, harmonic mean = 94), demonstrated that *transfer design* was identified as the strongest influence on learning in classroom training (r = .45, p < .001), and *transfer effort–performance expectations* was identified as the second strongest influence on learning (r = .27, p = .026). All other predictors exhibited non-significant paths (all other ps ≥ .22).

Classroom training and transfer. Regarding transfer as an outcome, six out of 12 meta-analytic correlations between transfer factors and transfer reached statistical significance (all ps ≤ .032) and were homogeneous (all ps ≥ .28 for Q statistics; for detailed statistical coefficients, see Table 33). Overall, meta-analytic correlations were larger for transfer compared to learning in classroom training. A large effect was revealed for *learner readiness*, medium effects for *transfer effort–performance expectations*, *performance coaching*, *motivation to transfer*, and *peer support*, whereas a small effect emerged for *openness to change*. Results of the MASEM (Model 4, harmonic mean = 87) demonstrated that *learner readiness* emerged as the most important driver for transfer in classroom training (r = .51, p < .001). In accordance to learning as outcome in classroom training, *transfer effort–performance expectations* showed an impact for transfer, too (r = .24, p = .011). *Performance coaching* was further identified as an influence on transfer (r = .25, p = .024), whereas all other predictors had non-significant paths to transfer (all other ps ≥ .059).

Learning as a mediational link between transfer factors and transfer

To answer the third research question, namely, which transfer factors affect transfer directly and/or indirectly, learning was explored as a mediator of the relationship between transfer factors and transfer, separately for e-learning and classroom training settings. Results from the previous saturated path models served as input information for further analyses in MASEM.

E-learning. To specify the model, the paths from *motivation to transfer* to learning and transfer, from *transfer design* to learning and transfer, from *learner readiness* to transfer, and from learning to transfer were contained. Conducting MASEM with learning as a mediator (Model 5, harmonic mean = 475), the overall model fit was satisfactory (χ^2 (1) = 6.86, RMSEA = .11, SRMR = .02, TLI = .93, CFI = .99). Even though the RMSEA is slightly above the recommended cut-off criteria of .08 (Vandenberg & Lance, 2000), it has been argued that the RMSEA can be a misleading fit indicator when the degrees of freedom in the model are small (Kenny, Kaniskan, & McCoach, 2014). As the presented model has only 1 degree of freedom, it can be regarded as acceptable. Direct and indirect effects of the model with its path coefficients are presented in Figure 25. *Motivation to transfer* showed the largest direct and indirect effects on transfer. *Transfer design* showed also direct and indirect effects on transfer, whereas *learner readiness* was not significant.

Classroom training. To specify the model, the paths from *transfer design* to learning, from *transfer effort–performance expectations* to learning and transfer, from *learner readiness* to transfer, from *performance coaching* to transfer, and from learning to transfer were contained. Conducting MASEM with learning as a mediator (Model 6a, harmonic mean = 90), the overall model fit was excellent (χ^2 (3) = 2.33, RMSEA = .00, SRMR = .03, TLI = 1.02, CFI = 1.00). As the direct effect of *transfer effort–performance expectations* on transfer was not significant (p = .20), this path was excluded, and a modified model was calculated (Model 6b, harmonic mean = 90). Again, the model resulted in an excellent model fit (χ^2 (4) = 3.97, RMSEA = .00, SRMR = .04, TLI = 1.00, CFI = 1.00). All path coefficients of the model (r = .15 to r = .48; all ps ≤ .035) and both indirect effects (*transfer design* on transfer via learning as mediator: r = .44, p < .001; *transfer effort–performance expectations* on transfer via learning as mediator: $r = .21$, $p = .027$) were substantial (see Figure 26) with *transfer design* showing a large indirect effect on transfer.

Comparing the strength of meta-analytic correlations across training settings

To compare e-learning and classroom training settings, for each of the two investigated training outcomes (learning and transfer), meta-analytic correlations of transfer factors that did not differ across training settings are described first. Second, transfer factors with larger meta-analytic correlations in e-learning compared to classroom training settings are presented. Third, transfer factors with larger meta-analytic correlations in classroom training compared to e-learning settings are briefly highlighted. Significance levels comparing the strength of meta-analytic correlations across training settings are presented in Table 33.

Transfer factors and learning. Results from analyses comparing the strength of the meta-analytic correlations from indepented samples revealed no differences across training settings for almost all transfer factors and learning (all ps ≥ .088). Descriptively, meta-analytic correlations were generally larger in e-learning compared to classroom training settings (except for *performance self-efficacy* and *performance coaching*). Only m*otivation to transfer* ($p < .001$) and *perceived content validity* ($p = .029$) showed significant differences across training settings, with larger meta-analytic correlations in e-learning settings. *Motivation to transfer* was a significant driver of learning only in e-learning settings but exhibited a non-significant path in the saturated classroom training setting model, confirming the result above.

Transfer factors and transfer. Again, almost all meta-analytic correlations between transfer factors and transfer did not show differences across training settings (all ps ≥ .056, except the ones presented next) with descriptively larger meta-analytic correlations primarily in e-learning compared to classroom training settings. Like for learning as an outcome, *motivation to transfer* ($p = .002$) and *perceived content validity* ($p = .012$) showed a significantly stronger influence in e-learning compared to classroom training settings. However, the meta-analytic correlation between *performance coaching* and transfer was significantly larger in classroom compared to e-learning settings ($p = .011$). This result was reflected in the mediation model (Model 6b) as *performance coaching* showed a substantial direct effect on transfer only in classroom training settings.

Table 33. *Overview over the Meta-Analytic Correlations with Level of Significance, Associated Number of Samples, Total Sample Sizes, 95% Confidence Intervals, Homogeneity Index Q with Degree of Freedom and Level of Significance Separately for E-learning and Classroom Training Settings*

Study Relationship	E-learning								Classroom Training									
	k	*N*	*ES r*	*p*	95% CI	*Q*	*df*	*p*	*k*	*N*	*ES r*	*p*	95% CI	*Q*	*df*	*p*	Model	*p**
Learning-Transfer	9	798	.669	.001	[.55;.79]	17.69	8	.024	3	176	.625	.001	[.47;-.78]	0.58	2	.748	5,6a,6b	.183
Learning-Lr	5	329	.320	.001	[.21;.43]	2.93	4	.570	2	72	.259	.101	[-.05;.57]	1.57	1	.210	1	.308
Learning-Motiv	7	697	.594	.001	[.52;.67]	4.33	6	.632	2	72	.258	.036	[.02;.50]	0.00	1	.989	1,3,5	.001
Learning-Cap	7	697	.232	.003	[.08;.38]	18.36	6	.005	2	72	.203	.099	[-.04;.44]	0.22	1	.640		.405
Learning-Peer	5	539	.296	.001	[.15;.44]	8.14	4	.087	2	72	.131	.335	[-.14;.40]	1.19	1	.276	1	.088
Learning-Ssup	6	658	.265	.001	[.13;.40]	11.42	5	.044	2	72	.128	.480	[-.23;.48]	2.05	1	.152		.130
Learning-Vali	7	697	.355	.001	[.27;.44]	6.79	6	.340	2	72	.131	.289	[-.11;.37]	0.12	1	.733	1	.029
Learning-Desg	7	697	.505	.001	[.38;.63]	11.69	6	.069	2	72	.496	.001	[.25;.74]	0.03	1	.860	1,3,5,6a,6b	.462
Learning-Pex	9	798	.362	.001	[.26;.47]	13.80	8	.087	3	176	.335	.001	[.18;.49]	0.37	2	.829	1,3,6a,6b	.357
Learning-Oue	4	286	.230	.092	[-.04;.50]	14.10	3	.003	2	72	.172	.269	[-.13;.48]	1.53	1	.215		.326
Learning-Ope	9	798	.198	.003	[.07;.33]	20.78	8	.008	3	176	.092	.567	[-.22;.41]	6.98	2	.031		.098
Learning-Effi	9	798	.167	.001	[.10;.24]	7.03	8	.533	3	176	.188	.032	[.02;.36]	2.35	2	.309	1,3	.398
Learning-Feed	8	759	.176	.020	[.03;.32]	23.14	7	.002	3	176	.248	.001[a]	[.10;.40]	1.80	2	.406	3	.185
Transfer-Lr	5	329	.414	.001	[.30;.52]	2.08	4	.721	2	72	.572	.001	[.33;.81]	0.36	1	.547	2,4,5,6a,6b	.056
Transfer-Motiv	7	697	.586	.001	[.49;.68]	7.83	6	.251	2	72	.305	.022	[.04;.57]	1.15	1	.284	2,4,5	.001
Transfer-Cap	7	697	.250	.022	[.04;.46]	37.92	6	.001	2	72	.267	.227	[-.17;.70]	3.01	1	.083		.443
Transfer-Peer	5	539	.274	.019	[.04;.50]	19.55	4	.001	2	72	.298	.016	[.06;.54]	0.07	1	.797	4	.419
Transfer-Ssup	6	658	.190	.031	[.02;.36]	18.99	5	.002	2	72	.072	.556	[-.17;.31]	0.66	1	.416		.171
Transfer-Vali	7	697	.451	.001	[.35;.55]	8.40	6	.210	2	72	.199	.435	[.30;.70]	3.97	1	.046	2	.012
Transfer-Desg	7	697	.535	.001	[.45;.62]	6.38	6	.382	2	72	.422	.187	[-.21;1.05]	6.28	1	.012	2,5	.122
Transfer-Pex	9	798	.314	.001	[.18;.45]	23.87	8	.002	3	176	.392	.001	[.24;.54]	1.22	2	.542	4,6a	.144
Transfer-Oue	4	286	.165	.086	[-.02;.35]	7.02	3	.071	2	72	.069	.576	[-.17;.31]	0.17	1	.682		.234
Transfer-Ope	9	798	.156	.027	[.02;.29]	24.17	8	.002	3	176	.166	.032	[.01;.32]	0.38	2	.826	4	.451
Transfer-Effi	9	798	.139	.010	[.03;.25]	14.28	8	.075	3	176	.216	.103	[-.04;.47]	4.79	2	.091	2	.172

Transfer-Feed	8	759	.194	.047	[.00;.10]	39.52	7	.001	3	176	.372	.001	[.22;.52]	0.76	2	.685	4,6a,6b	.011
Lr-Motiv	5	329	.448	.001	[.30;.59]	6.47	4	.167	2	72	.149	.530	[-.32;.62]	3.48	1	.062	1,2,3,4	
Lr-Cap									2	72	.459	.001	[.22;.70]	0.03	1	.854	3	
Lr-Peer	4	210	.266	.061	[-.01;.54]	11.56	3	.009	2	72	.309	.012	[.07;.55]	0.01	1	.904	1,4	
Lr-Vali	5	329	.490	.001	[.38;.60]	3.49	4	.480									1,2	
Lr-Desg	5	329	.466	.001	[.36;.58]	0.90	4	.925	2	72	.262	.583	[-.67;1.20]	13.96	1	.001	1,2,3	
Lr-Pex	5	329	.322	.001	[.21;.43]	3.76	4	.440	2	72	.237	.054	[.00;.48]	0.02	1	.888	1,3,4	
Lr-Ope									2	72	.328	.047	[.00;.65]	1.71	1	.191	4	
Lr-Effi	5	329	.170	.003	[.06;.28]	2.74	4	.603	2	72	.084	.646	[-.27;.44]	2.07	1	.150	1,2,3,4	
Lr-Feed									2	72	.255	.039	[.01;.50]	0.21	1	.648	3, 4	
Motiv-Cap									2	72	.291	.018	[.05;.53]	0.98	1	.323	3	
Motiv-Peer	5	539	.469	.001	[.25;.69]	18.69	4	.001	2	72	.365	.003	[.12;.61]	0.27	1	.606	1,4	
Motiv-Vali	7	697	.440	.001	[.28;.60]	21.55	6	.002									1,2	
Motiv-Desg	7	697	.520	.001	[37;.67]	18.15	6	.006	2	72	.499	.001	[.26;.74]	0.39	1	.532	1,2,3	
Motiv-Pex	7	697	.449	.001	[.28;.62]	23.83	6	.001	2	72	.325	.072	[-.03;.68]	2.03	1	.154	1,3,4	
Motiv-Ope									2	72	.240	.051	[.00;.48]	0.91	1	.340	4	
Motiv-Effi	7	697	.146	.087	[-.02;.31]	22.51	6	.001[a]	2	72	.375	.002	[.13;.62]	0.91	1	.340	1,2,3,4	
Motiv-Feed									2	72	.292	.018	[.05;.53]	0.33	1	.568	3,4	
Cap-Peer									2	72	.549	.001[a]	[.22;.88]	1.78	1	.183		
Cap-Desg									2	72	.245	.458	[-.40;.89]	6.65	1	.010	3	
Cap-Pex									2	72	.533	.001	[.29;.77]	0.91	1	.340	3	
Cap-Ope									2	72	.440	.001	[.20;.68]	0.95	1	.329		
Cap-Effi									2	72	.521	.001	[.28;.76]	0.62	1	.430	3	
Cap-Feed									2	72	.464	.001	[.21;.72]	1.10	1	.295	3	
Peer-Vali	5	539	.412	.001	[.29;.54]	6.03	4	.197									1	
Peer-Desg	5	539	.449	.001	[.30;.59]	7.79	4	.100	2	72	.261	.034	[.02;.50]	0.10	1	.752	1	
Peer-Pex	5	539	.444	.001	[.29;.60]	8.78	4	.067	2	72	.539	.001	[.23;.85]	1.54	1	.214	1,4	
Peer-Ope									2	72	.591	.002	[.22;.96]	2.19	1	.139	4	
Peer-Effi	5	539	.234	.004	[.08;.39]	9.32	4	.054	2	72	.380	.002	[.14;.62]	0.04	1	.848	1,4	

Peer-Feed									2	72	.702	.001	[.42;.99]	1.37	1	.243	4
Vali-Desg	7	697	.665	.001	[.59;.74]	5.34	6	.501									1,2
Vali-Pex	7	697	.272	.006	[.08;.47]	30.94	6	.001									1
Vali-Effi	7	697	.124	.156	[-.05;.29]	23.78	6	.001									1,2
Desg-Pex	7	697	.444	.001	[.29;.60]	20.26	6	.003	2	72	.296	.016	[.05;.54]	0.07	1	.793	1,3
Desg-Ope									2	72	-.011	.928	[-.25;.23]	0.86	1	.354	
Desg-Effi	7	697	.278	.001	[.13;.42]	17.06	6	.009	2	72	.298	.019	[.05;.55]	1.05	1	.306	1,2,3
Desg-Feed									2	72	.206	.094	[-.04;.45]	0.20	1	.653	3
Pex-Ope									3	176	.314	.001	[.16;.47]	1.89	2	.389	4
Pex-Effi	9	798	.410	.001	[.27;.55]	26.66	8	.001	3	176	.676	.001	[.52;.83]	0.34	2	.845	1,3,4
Pex-Feed									3	176	.388	.001	[.20;.58]	2.79	2	.248	3,4
Ope-Effi									3	176	.360	.001	[.21;.51]	0.41	2	.817	4
Ope-Feed									3	176	.349	.041	[.02;.68]	7.88	2	.020	4
Effi-Feed									3	176	.279	.001	[.13;.43]	0.60	2	.741	3,4

Note. Model 1 refers to transfer factors on learning in e-learning settings. Model 2 refers to transfer factors on transfer in e-learning settings. Model 3 refers to transfer factors on learning in classroom training settings. Model 4 refers to transfer factors on transfer in classroom training settings. p^* = significance level comparing meta-analytic correlations (*ES r*) between e-learning and classroom training settings.

[a] = the exact *p*-value of $p = .001$ was achieved, all other presented *p*-values $p < .001$.

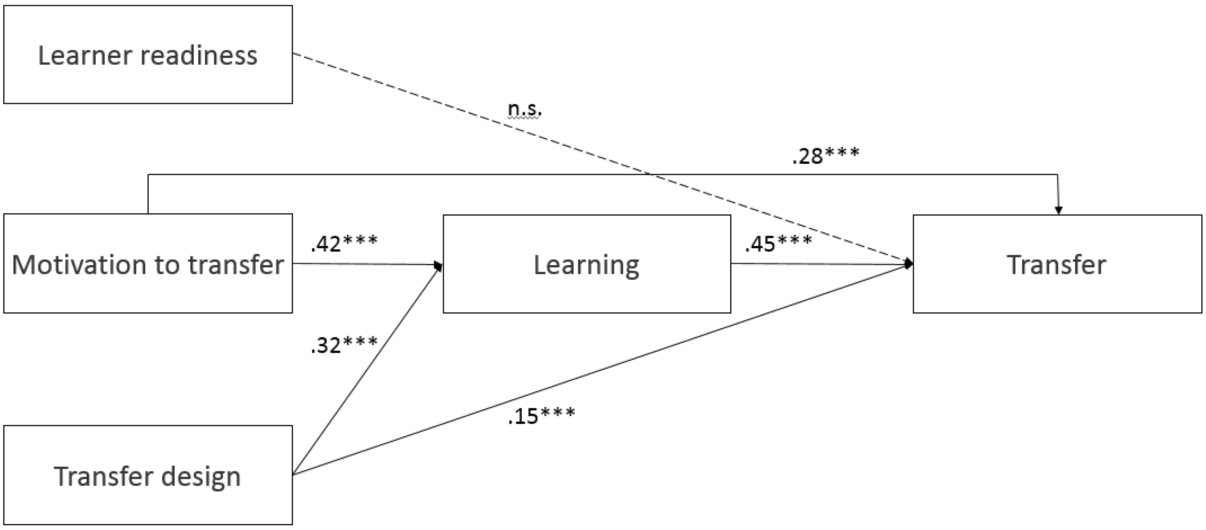

Figure 25. Results of the path model of the effects of transfer factors on transfer via learning as a mediator in e-learning settings (Model 5, harmonic mean = 475). The overall model fit was satisfactory ($\chi2$ [1] = 6.86, RMSEA = .11, SRMR = .02, TLI = .93, CFI = .99). *** $p < .001$.

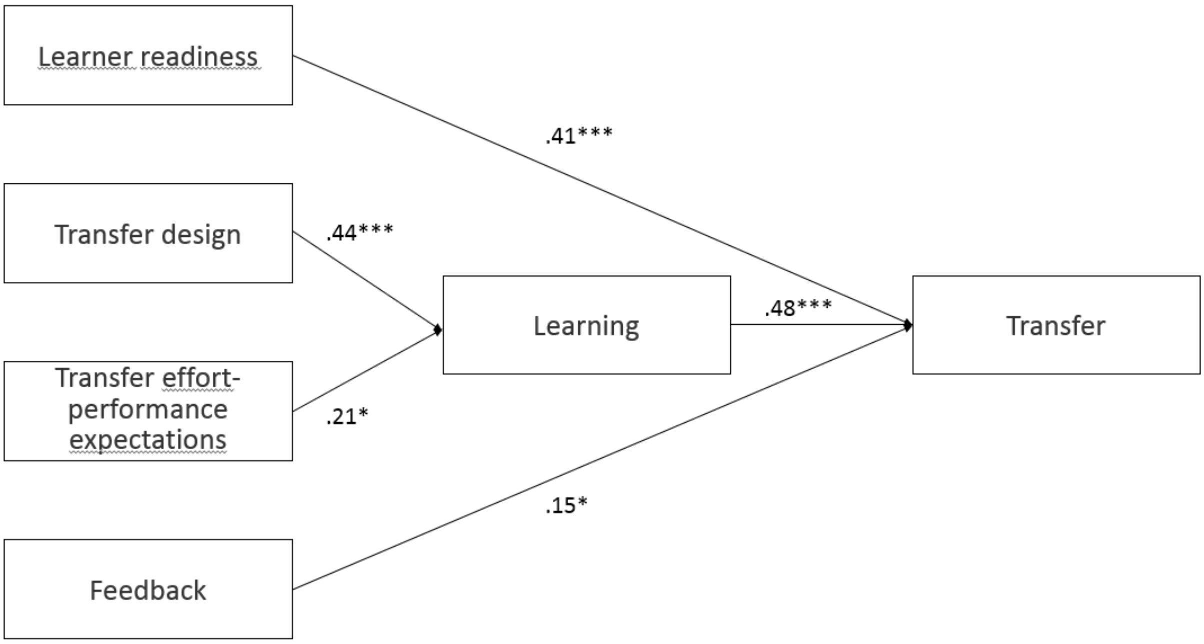

Figure 26. Results of the path model of the effects of transfer factors on transfer via learning as a mediator in classroom training settings (Model 6b, harmonic mean = 90). The overall model fit was excellent ($\chi2$ [4] = 3.97, RMSEA = .00, SRMR = .04, TLI = 1.00, CFI = 1.00). * $p < .05$. *** $p < .001$.

5.3.4 Discussion

Summary

The aim of the study was to shed light on the transfer problem in corporate e-learning settings. Hence, this meta-analysis comprising 12 corporate samples exploratively investigated which transfer factors are important for success of learning and success of transfer, respectively. The study explored transfer factors that help ensure training success especially in corporate e-learning settings.

First, higher meta-analytic correlations between transfer factors and training success measures were generally found for success of transfer compared to success of learning in both training settings. For example, *learner readiness* was not substantially related to success of learning but it was to success of transfer in both training settings, as confirmed in the saturated path models (Models 1, 2, 3, and 4). Different explanations are conceivable: (a) success of learning (measured by the Q4TE) is more closely related to the training situation and (b) transfer factors (measured by the GLTSI) refer especially to the transfer situation. As the GLTSI is constructed to measure factors that affect training transfer, this effect is in the nature of the questionnaire. According to this, item wordings of GLTSI factors rather point towards transfer than towards learning (e.g., regarding the scale *learner readiness*: "Before the training I had a good understanding of how it would fit my job-related development" [Holton et al., 2000, pp. 344-346]). Further, to successfully transfer knowledge, skills, and attitudes gained in training to the workplace (Baldwin & Ford, 1988), the transfer process requires a longer time period and is more complex compared to a successful learning process. Thus, various factors with higher correlations affect training transfer which was confirmed in the study. These results imply that transfer factors are potentially more important for success of transfer.

Second, I turn to answer our questions (a) which transfer factors are important for success of learning, (b) which ones are important for success of transfer, and (c) which transfer factors indirectly affect success of transfer via success of learning and which transfer factors affect it directly. E-learning settings are discussed first and classroom training settings later.

In e-learning settings, *motivation to transfer* and *transfer design* showed large meta-analytic correlations on success of learning and also on success of transfer which is also reflected in substantial direct effects in the saturated path models.

Indicating consistent findings with reference to success of learning in e-learning settings, *motivation to transfer* had the strongest influence on success of transfer, and *transfer design* showed the second strongest influence (Models 1 and 2). Further, *learner readiness* was identified as relevant to success of transfer in e-learning settings (Model 2), but the path coefficient was not significant in the mediation model for e-learning settings.

Our result that *motivation to transfer* was identified as the most important key factor to ensure success of transfer is consistent with findings, for example, by Machin & Fogarty (1997) and more recently by Grohmann et al. (2014). Our results underline Gegenfurtner et al.'s (2009, p. 403) statement that "motivation is essential for training transfer".

Further, it is especially important for success of transfer that "the degree to which training has been designed and delivered" and also the degree to which "training instructions match job requirements" (*transfer design*) are very high in e-learning settings to give trainees the ability to transfer learning on the job (Holton et al., 2000, pp. 344-346).

Summarizing results for e-learning settings, *motivation to transfer* as well as *transfer design* showed substantial direct and indirect effects in the mediation model with success of learning as a mediator. Finding a strong and robust model, this speaks in favor of both the validity and generalizability of the presented results.

Focusing on classroom training settings, similar influencing transfer factors were found in the saturated path models assessing direct effects on success of learning and success of transfer (Models 3 and 4).

Similar to e-learning, *transfer design* showed a large effect on success of learning (Model 3) indicated by the the strongest path coefficient in this model. Unlike in e-learning, the meta-analytic correlation between *transfer design* and success of transfer was neither significant nor homogeneous which was due to high variance across samples and thus was not included in the saturated path model. In the mediation model, *transfer design* showed the largest indirect effect. These results indicate that designing a transfer-supportive training setting that matches job requirements is very important in classroom training, too.

Likewise in e-learning, *learner readiness* was substantial for success of transfer. This was indicated by (a) the highest path coefficient on success of transfer in the saturated model

(Model 4) and (b) by a large direct effect on success of transfer in the mediation model. These results show that trainees who are prepared to enter and participate in training are more likely to transfer successfully.

Unlike in e-learning, the transfer factor *transfer effort–performance expectations* was important for success of learning and success of transfer in classroom training. This transfer factor showed the second highest path coefficient on success of learning in the saturated model (Model 3) and a significant indirect effect on success of transfer via success of learning as a mediator. Even though no differences occurred across training settings when comparing meta-analytic correlations, *transfer effort–performance expectations* significantly contributed to training success only in classroom training settings. This effect can be caused by the fact that other transfer factors contributed considerably more to success of transfer in e-learning (i.e., *motivation to transfer*). An explanation might be that e-learning trainees are not that aware of the connection between transfer effort and changes in job performance. They rather have to extrapolate this link on their own which develops across time. Whereas in classroom training, this connection is more obvious and "visible" as (a) trainees rather question the trainer during training and/or (b) the trainer illustrates how the content learned applies to everyday work.

Unlike in e-learning, *performance coaching* was further important for success of transfer in classroom training. This was indicated by (a) the second strongest path coefficient on success of transfer in the saturated model (Model 4) and (b) by a small direct effect on success of transfer in the mediation model. Hence, in classroom training, it seems important that the organization formally and informally feedbacks trainee's job performance (Holton et al., 2000, pp. 344-346).

For the sake of completeness at this point but it is discussed later, unlike in e-learning, *motivation to transfer* was not significant in the transfer process in classroom training.

Third, when comparing the strength of meta-analytic correlations across training settings, the majority of meta-analytic correlations did not differ statistically across training settings. From a descriptive perspective, most of the meta-analytic correlations were smaller in classroom compared to e-learning settings. Significantly higher correlations were only found for *motivation to transfer* and *perceived content validity (*regarding both success of learning and success of transfer) in e-learning settings (discussed first), whereas significantly higher correlations were only found for *performance coaching* and success of transfer in classroom training settings (discussed second).

Even though *motivation to transfer* showed significant and homogeneous meta-analytic correlations and consequently was considered in the saturated path models, the path coefficients were non-significant in MASEM in classroom trainings. The reason for this effect is mainly due to the samples included in the meta-analysis. In our study, the two classroom samples were indifferent regarding the correlation between *motivation to transfer* and success of transfer (sample 1: r = .14; sample 2: r = .41). As sample 1 reduced the meta-analytic correlation to ES r = .31 (p = .022) and a small sample size *(n* = 72) was given, consequently, the chances to be a relevant predictor of training transfer in MASEM were reduced. Further, for learning and transferring successfully in classroom training, *motivation to transfer* might not be derived from the trainee's own accord; instead, it might be triggered in the group setting from the trainer and/or other trainees (extrinsic motivation). Contrarily, in e-learning settings, intrinsic motivation is rather triggered as more self-regulated learning is required from trainees in both learning and subsequent transfer processes. Research literature demonstrated that trainees driven by intrinsic motivation to participate in training showed higher motivation in the course of training itself and also higher motivation to learn, both resulting in higher success of learning and success of transfer, while extrinsic motivation did not contribute significantly to success of learning and success of transfer (Facteau et al., 1995). In line with these findings, intrinsic motivation (r = .34) has a greater impact on remembering the content learned than extrinsic motivation (r = .05; Kontoghiorghes, 2001). However, transfer of learning is greatest when trainees were additionally motivated by extrinsic components that are anchored in their field of work (Taylor et al., 2005). Thus, high levels of *motivation to transfer* are essential to successfully learn and apply training contents to the workplace. The same applies to *perceived content validity* in e-learning settings. That is, trainees who perceive high levels of consistency between training and job content are more likely to learn and transfer successfully in e-learning, whereas in classroom training, this relation is not that closely connected.

The meta-analytic correlation between *performance coaching* and success of transfer was stronger in classroom training settings. *Performance coaching* seems especially important in classroom training because trainees are more used to constant feedback during training provided by the trainer and/or by other trainees, whereas in e-learning, less or no feedback is provided during training. However, in the transfer process (workplace) itself, feedback, for example, from supervisors and/or colleagues does not differ across training settings (i.e., equal environments). Consequently, classroom trainees who are used to, for example, constant positive feedback more likely need this positive reinforcement to transfer successfully. Overall, it is assumed that success of transfer depends very much on the baseline level of feedback provided during training.

Additionally, even though meta-analytic correlations did not statistically differ across training settings (p = .056), the direct effect of *learner readiness* on success of transfer was only significant in classroom training settings. It seems more obvious and easier for classroom trainees to question, for example, aims or consequences of the training (e.g., by "bombarding" the trainer with questions who adds clarity) and thus to satisfy trainee's expectations. Therefore, the influence of *learner readiness* on training transfer might be greater in classroom compared to e-learning settings.

Summarizing results for the comparison across training settings, the majority of transfer factors did not differ between e-learning and classroom training settings. Thus, it can be concluded that the training setting is not excessively relevant. Rather, the specific factors contribute to success of learning and success of transfer, respectively. Besides *transfer design* which was substantial in both training settings as confirmed in high meta-analytic correlations and in the mediation models, results identified *motivation to transfer* (Gegenfurtner et al., 2009) as the most important key to ensure success of learning and success of transfer in corporate e-learning settings.

Limitations

This study was conducted to shed some light on transfer factors important for the growing presence of e-learning settings. In this regard, transfer factors, which have been less studied in e-learning and rather well—investigated in classroom training settings, have been applied to e-learning contexts and have been tested exploratively using the meta-analytic approach. However, after identifying the most important transfer factors, structural equation models with less factors influencing training transfer were tested in the end. Further, investigating the comparison across training settings, greater sample sizes in classroom training settings seem essential. Nevertheless, the study aimed at shedding light on the transfer problem specifically in corporate e-learning settings, and the comparison with classroom settings was out of the main scope of this study. However, the comparison of the strength of meta-analytic correlations across training settings was additionally explored in order to guide future research which should look specifically at differences between e-learning and classroom training settings.

The study is further limited as the samples of included studies were rather small, and the rated trainings were very heterogeneous in terms of training settings, training contents, and training duration. Even though this heterogeneity in samples can be seen as advantageous, further studies are needed to investigate transfer factors in corporate e-learnng settings. Although one advantage of the study is that not a single organization was investigated, collecting data from other researchers and comprising (a) published studies from the research literature and (b) other companies with further branches of industry would not only increase the sample size but also enlarge the impact of this explorative study.

Self-report data is another limitation. Other research studies on transfer factors also referred to self-reportings as it is a rather common method to assess training transfer. On the one hand, assessing objective training success measures such as concrete performance results of each trainee seems more promising. On the other hand, collecting this amount of data in a corporate environment is very complex and time-consuming. Yet, the previous study (Study 5) demonstrated that trainees are very much capable of evaluating their training success subjectively. Nonetheless, assessing both objective and subjective training success measures seem more profound. It would be desirable for future research to enrich the self-reported findings in this study by objective criteria or third-party ratings of training success (e.g., concrete performance results of each trainee).

Implications for practice

As this meta-analysis identified important drivers of successful training transfer, it contributes to HRD practice especially for e-learning settings, thereby, it speaks to the highly relevant transfer problem claimed by Baldwin and Ford (1988) and more recently by Saks et al. (2014). Accurately diagnosing strengths and weaknesses of transfer processes creates opportunities for performance improvement (Chen et al., 2005). If performance of employees improves, the performance of the organization is likely to improve as well, which stimulates the company's economic growth. Leverage points for practical implications in the corporate environment refer to the four sets of factors according to Holton et al. (2000): motivation, ability, work environment, and secondary influences which are also known as trainee characteristics.

Our meta-analysis showed that *motivation to transfer*, a variable on the individual's level according to Baldwin and Ford's model (1988), played the biggest role in predicting success of learning and success of transfer in corporate e-learning. To transfer training successfully, trainees need to achieve high levels regarding "the direction, intensity and persistence of effort toward utilizing in a work setting skills and knowledge learned" (Holton et al., 2000, pp. 344-346). Conceivably, trainers need to be encouraged to incorporated specific motivational factors in learning processes during training, depending on training content, branch of industry, or trainees. Regarding further motivational issues, particularly before training, T&D professionals should raise the awareness in trainees that effort devoted to transferring learning will lead to changes in job performance. For example, learning content and examples should be adapted to everyday work situations.

In the ability domain, *designing* the training transfer-supportively and making sure that training instructions match job requirements seems essential in both training settings. The trainer's task is to provide such support, which seems to be even more crucial in e-learning settings. For example, a virtual trainer can provide this guidance for trainees.

As to secondary influences (trainee characteristics), it is important that T&D professionals prepare trainees to actively participate in training, since secondary influences are perceived to affect motivation and then further to affect individual performance. For example, trainers and trainees themselves should raise "the extent to which individuals are prepared to enter and participate in training" (Holton et al., 2000, pp. 344-346) to the highes level possible.

Regarding the work environment and training in general, providing feedback from supervisors after training enhances success of transfer especially in classroom training.

In sum, results emphasize the importance of promoting variable, training-specific transfer factors in corporate training and transfer environments. In particular, motivation to transfer needs to be facilitated so that time, money, and resources of an organization are not wasted (Latham, 2007, p. 3).

CHAPTER 6 CONCLUSION OF CAREER CONSTRUCTION ACROSS THE LIFE SPAN

This doctoral thesis contributes to deepen our understanding of constructs that play a key role in individuals' vocational career construction. In this regard, many previous studies have focused exclusively on a specific phase of an individual's career. However, post-modern societies require continuous investments in one's career to adapt to changing environments throughout the life span. Consequently, this dissertation takes a broad approach to capture a wide spectrum of career construction processes.

According to Super's (1990) developmental stage framework, individuals have to manage vocational developmental tasks corresponding to each of the developmental life stages to be career mature across the life span. As the two stages exploration and maintenance direct individuals' future career pathways, they are especially important in individuals' vocational career construction. Therefore, both of them are addressed in this dissertation.

In the exploration stage, individuals have to make vocational choices in order to strive for career direction. Such vocational choices are guided by vocational interests in this early career stage, which provide the basis for individuals' career construction in the following career stages. In the maintenance stage, on the contrary, individuals must strive for career adaptability to respond to rapidly changing labor needs of post-modern societies once an occupation is chosen to maintain career success. Therefore, individuals focus on an active development of their individual careers. A crucial building block of career development is acquiring new knowledge, skills, or abilities through the successful completion of trainings. In this regard, guaranteeing training success is important to ensure that individuals make satisfactory progress towards their career development.

Noteworthy, even though Super (1990) has arranged these different phases in a stage model along the adult development, the stages blend into each other in post-modern societies. Thus, transitions are smoother across stages and not limited to a specific age range. All of the six presented empirical studies have been conducted with the overarching goal in mind to shed more light on constructs relevant to individuals' vocational career construction processes in early and later career stages, respectively. Beyond the results and implications previously presented within each study's horizon, this chapter explores how these results are interrelated on a broader scale. First, the main outcomes and findings of the diverse studies are briefly summarized and discussed to answer how vocational interests shape career choice processes in early career stages and how trainings can be designed to be useful building blocks of successful career development in later career stages. Second, implications for research and practitioners are presented and strengths and weaknesses of the research conducted within this dissertation are outlined. Finally, directions for future research are illustrated, before closing up with some concluding remarks.

6.1 Overview of insights on career construction processes

6.1.1 Early career stages: from vocational interests to career choice

As outlined above, early career choice processes are closely linked to vocational interests. In this regard, the current dissertation first meta-analytically confirms the relationship between vocational interests and cognitive abilities as both constructs are key factors for choosing a career. In a second step, vocational interest inventories are studied in more detail to better understand how test fairness can be best established for a test assessing such basic, yet career-pointing interests. Taken together, the specific findings of the three empirical studies help to advance vocational interest research. An overview of the main research questions, methods, and results are provided in Table 34. In the following, answers to a set of current questions regarding vocational interests as career drivers are outlined and discussed across studies.

How are vocational interests related to specific cognitive abilities?

The first study of this dissertation shed meta-analytic light on the relationships between vocational interests and cognitive abilities. Providing a comprehensive quantitative summary, the main results are in accordance with Ackerman and Heggestad's (1997) review that Realistic (with mechanical knowledge, spatial -, and numerical abilities), Investigative (with *g*, induction, numerical -, spatial -, and verbal abilities), and Artistic interests (with verbal abilities) were positively linked to cognitive abilities. In contrast, negligible or negative correlations were found for Social (with mechanical knowledge, verbal abilities) and Enterprising (with verbal abilities) interests and specific cognitive abilities. It might be the case that occupations that strongly emphasize dealing with social and economic relations do not require high levels of cognitive abilities but at least average levels of *g*, verbal, and numerical abilities (L. S. Gottfredson, 1986). Furthermore, in contrast to previous findings (e.g., Carless, 1999; Reeve & Heggestad, 2004), gender was not found to be a substantial moderator of the relationship between vocational interests and cognitive abilities. Similar results were found for interests and personality (Staggs et al., 2007). For specific cognitive abilities that are influenced by experience and knowledge acquisition according to the CHC framework (i.e., language development and quantitative knowledge; Schneider & McGrew, 2012), stronger relationships were found for older samples than for younger samples. As (a) according to cognitive investment theories (Cattell, 1987; Schmidt, 2011) interests and personality guide the development of crystallized intelligence, and (b) general interest in knowledge acquisition correlates positively with crystallized intelligence, acquired knowledge, and academic performance (Von Stumm & Ackerman, 2013), the importance of interests within developmental processes is obvious. Taken together, assessing vocational interests in addition to cognitive ability measures provides further information that is crucial for the individual's vocational career choice that, in the end, might lead to high levels of career success.

As such, vocational interests are well-known predictors of educational choice (Hansen & Neuman, 1999), degree completion (Webb et al., 2002), occupational choice (Hansen & Dik, 2005), and occupational satisfaction (Tsabari et al., 2005), which has been reinforced by the results of the presented meta-analysis. Since important educational and vocational choices can be based on results of interest inventories, it is immensely important to adequately measure vocational interests with the lowest measurement bias possible. Therefore, the second and third study are dedicated to assess test fairness of vocational interests across genders in order to take measures to correct for the bias, where necessary.

Guaranteeing test fairness across gender: Can interest inventories eliminate gender differences in instrument validity and prediction?

In line with meta-analytical results, findings of Studies 2 and 3 confirmed considerable gender (mean) differences in a standard RIASEC interest inventory, a frequently used inventory to assess vocational interests. More specifically, women tend to report stronger Social, Artistic, and Conventional interests than men, whereas men are more likely than women to prefer Realistic and Investigative activities (Lippa, 1998; Su et al., 2009). Indeed, previous findings suggest that gender differences in vocational interests are among the largest differences in the field of individual differences (Lubinski, 2000) and in the psychological domain in general (Hyde, 2005).

Albeit possessing the same underlying trait level, women and men respond differently to certain interest items. The question of where these large gender-related mean differences within the vocational interest domain originate from leads to the issue of measurement bias as they might be interpreted as an indicator of unfairness in interest inventories. Ever since the 1970s, various approaches to how these gender-specific mean differences might be reduced were discussed and respective guidelines were formulated (e.g., Guidelines for the Assessment for Sex Bias and Sex Fairness in Career Interest Inventories; NIE, 1975) to ensure that diverse aspects of gender fairness are considered when developing interest inventories. One recommended, albeit criticized approach to ensure test fairness is to eliminate those items showing substantial gender differences (e.g., UNIACT-R [ACT Program, 1995]). However, studies reveal that this approach can affect the instrument's construct validity (Russell, 2007; Su et al., 2009), which goes against the general goal to develop gender-fair and valid instruments. Therefore, there seems to be a trade-off between optimizing an interest inventory in terms of gender differences and ensuring its construct validity.

Facing the long-time debate on consequences of eliminating items showing gender-specific DIF, the issues of differential validity and differential prediction using DIF analysis in the vocational interest domain were addressed for the first time (Study 2). The results showed that removing items showing large DIF (a) eliminated differential validity in the total sample (differential validity on the group level was reduced, yet not eliminated completely) and (b) eliminated prediction biases. Therefore, the instrument's predictive validity was not affected.

Beyond predictive validity, structural validity is crucial to instrument's construct validity. Study 3 confirmed invariance in the structure of a standard RIASEC interest inventory across gender (Darcy & Tracey, 2007). Moreover, the study highlighted that structural assumptions of the interest inventory remained invariant when items showing large gender-specific DIF were eliminated. Thus, controlling for DIF has been shown to be one very powerful way to reduce gender biases in interest inventories and to optimize gender fairness in the test development process (Wetzel et al., 2012). Interestingly, these results provide much needed empirically support for previously proposed test development standards (AERA, APA, & NCME, 1999; NIE, 1975). Of course, we cannot guarantee test fairness even with the applied corrections. Nevertheless, approaches to substantially reduce measurement bias and thus establishing or improving test fairness are confirmed in light of studies conducted in this dissertation.

Table 34. *Summary of the Aims, Methods, and Main Results of Empirical Studies on Vocational Interests Relevant for Career Choice in Early Career Stages*

Study	Aims	Method	Main results
1 (Section 3.1)	•Examining the nature and magnitude of the relationship between cognitive abilities and vocational interests.	•Meta-analysis of 27 studies with 29 independent samples and an overall sample size of 55,297 participants was conducted.	•The study demonstrated meaningful relations between cognitive abilities and vocational interests. •Meta-analytic coefficients ranged from -.29 to .47. •Coefficients' strength and direction were comparable for females and males. •Age and birth cohort were established as moderators of the relation between interests and cognitive abilities.
2 (Section 3.2)	•Addressing the issue of differential validity and differential prediction in the vocational interest domain as one major concern of test fairness (Following the suggestions of the National Institute of Education and American Psychological Association Standards). •Investigating potential prediction bias in vocational interest measures.	•Comparing gender-specific validity coefficients for the prediction of person-environment fit and satisfaction to test for differential validity •Examining gender differences in the slopes and intercepts of the regression model predicting person-environment fit to test for differential prediction. •*N* = 736 participants enrolled in vocational training and university programs.	•Results showed evidence of differential validity and some indications of differential prediction in a standard Holland interest inventory. •Removing items showing large gender-specific DIF, that is, controlling for measurement bias, slightly reduced prediction bias.
3 (Section 3.3)	•Investigating consequences of eliminating items showing gender-specific DIF on the psychometric structure of a standard RIASEC interest inventory.	•Holland's hexagonal model was tested for structural invariance using a confirmatory methodological approach (confirmatory factor analysis and randomization tests of hypothesized order relations). •*N* = 736 participants enrolled in vocational training and university programs.	•Results suggested that eliminating items showing gender-specific DIF had no considerable influence on the instrument's psychometric structure. •Considering DIF is one possibility to significantly improve test fairness when developing interest inventories.

6.1.2 Later career stages: from training evaluation to career development

As outlined above, later career development processes are closely linked to the successful completion of trainings. In this regard, the current dissertation first studied to what extent do corporate businesses apply evidence-based actions to maximize training effectiveness (theory-practice-transfer). In a second step, e-learning and classroom training settings were systematically compared (a) in terms of objective and subjective training success measures within an experimental study and (b) in terms of substantial transfer factors for training success. Taken together, the specific findings of the three empirical studies help to advance training evaluation research. An overview of the main research questions, methods, and results are provided in Table 35. In the following, answers to a set of current questions regarding training evaluation are outlined and discussed across studies.

Doing the reality check: To what extent do corporate businesses apply evidence-based actions to maximize training effectiveness?

To answer this question with a representative sample, Study 4 investigated T&D professionals of large-scale enterprises with high turnovers and a great number of employees which are among the best employers in Germany. In support of the increasing shift from traditional classroom training to e-learning, Study 4 identified that more than two-thirds of the companies investigated offered e-learning as part of their HRD activities, compared to one-third found in earlier studies conducted by the MMB (2010, 2012). This speaks to the investigation of assessing transfer support actions not only in classroom training but also in e-learning settings.

In sum, the study revealed that e-learning-specific actions for maximizing training effectiveness were widely implemented in corporate practice to varying degrees. For example, results indicated that e-learning settings were generally designed adequately. When considering the different time periods before, during, and after training, it became clear that transfer support actions were mostly implemented before and during training, both in e-learning as well as in classroom trainings. The finding that only very few transfer support actions were implemented after training confirms similar findings by Salas et al. (2012) and Van Buren and Erskine (2002). Further, results demonstrated that before training, low values were obtained for adapting the training environment to meet the needs of older workers in both training settings. This finding calls for a demand of ensuring the fit with trainees' needs particularly in terms of an aging process of employees.

Nevertheless, transfer support actions have been shown to play a crucial role in ensuring successful learning and application of training content to the real world setting, which is the ultimate goal of training. Therefore, the effect that only little means are taken after training contributes to the so-called transfer problem (e.g., Baldwin & Ford, 1988; Grossman & Salas, 2011; Saks et al., 2014), which refers to phenomenon that the knowledge and skills acquired in training are insufficiently transferred to the workplace. Hence, results lead to practical implications in corporate practice as these actions are most relevant to maximize training success on a long-term basis (Salas et al., 2012).

Following up on this urgent need and answering Aguinis and Kraiger's (2009) call for effectiveness studies examining predictors of success of learning and success of transfer in training, Study 5 and 6 explored measures to enhance such training success across different training settings (i.e., e-learning and classroom training). Specifically, Study 5 was conducted as an experimental study on condition of ensuring good transfer-supportive training settings (i.e., controlling for the same baseline and maintaining transfer support actions before and during training). Study 5 aimed at identifying differences in predictors of success of learning and success of transfer between e-learning and classroom training, whereas Study 6 aimed at identifying transfer factors which are substantial for training success specifically in corporate e-learning settings.

Which are the differences between e-learning and classroom training regarding training success?

Two different methodological approaches were used to shed light on this question regarding differences between e-learning and classroom training. The large retrospective survey in Study 4 revealed that fewer actions are taken to maximize training effectiveness in e-learning compared to classroom training settings before, during, and after training. The field experiment in Study 5 reacted to the fact that, so far, only few empirical studies have been conducted in corporate practice that systematically measure and compare differences in training success between e-learning and classroom training dealing with the same training contents. In both training settings, training success was measured objectively in terms of performance tests and subjectively by trainees at two points in time with a time interval of six to eight weeks. With respect to objective training success, factual knowledge was higher in an e-learning setting, whereas applied knowledge showed similar scores in e-learning and classroom training six to eight weeks after the training course, confirming results found by the MMB (2010) and Sitzmann et al. (2006). Differences in subjective training success (i.e. success of learning and success of transfer) were only detected right after the training, whereas six to eight weeks later, participants from both settings perceived their training as equally effective. In sum, findings of Study 5 indicated that in the end both training settings get similar results for applied knowledge, and that, consequently, none of the training settings dominates the other. However, regardless of the training setting, it is crucial to ensure both success of learning and success of transfer, as outlined above. As classroom training and e-learning differ significantly in their structural characteristics (e-learning was rather flexible as to time and place and not guided by a trainer, whereas traditional classroom training was fixed in time and in location, clearly structured, and guided by a trainer), it is important to better understand which transfer support actions help trainees to make the most out of their training in each setting. As classroom settings are widely explored by previous research, Study 6 emphasized transfer support actions in the largely underexplored e-learning setting and offers only preliminary comparisons to classroom settings.

What can T&D professionals actively do to support learning and transfer for effective trainings?

The implementation of e-learning settings is tremendously increasing as can be seen from the results of Study 4, which corroborates the prediction of the MMB (2012) that e-learning and blended learning as training settings will play a leading role in the future. Yet, Study 4 revealed that fewer transfer support actions are implemented in e-learning compared to classroom training settings, which implies a huge potential to improve training effectiveness especially in e-learning settings. In order to shed more light on training success processes in e-learning settings and to guide practitioners how to invest the rare resources (e.g., time, money, etc.) in transfer support activities, Study 6 explored transfer factors which are substantial for successfully transfer training in different subsamples varying in branches of industry and content learned. As a result, a transfer-supportive training design and a training-specific activation of trainees' motivation to transfer were identified as key determinants to ensure training success in corporate e-learning settings. Structural relationships with the two training outcome measures (success of learning and success of transfer) underline the importance of promoting variable predictors of transfer, especially intrinsic-generated motivation to transfer, in corporate e-learning environments and transfer environments to transfer successfully.

Table 35. *Summary of the Aims, Methods, and Main Results of Empirical Studies on Training Evaluation Relevant for Career Development in Later Career Stages*

Study	Aims	Method	Main results
4 (Section 5.1)	•To what extent do corporate businesses apply well-investigated evidence-based recommendations and best practices for maximizing training effectiveness? •Investigating differences between e-learning and classroom training.	•Online survey •Corporate environment; HRDs of the companies awarded as “best employers in Germany” •*N* = 134 •72% large-scale enterprises, various branches of industry (15), high turnovers, national or even international scope of application	•Before, during, and after training: high implementation rate (almost always) in both training settings. •Classroom training dominates e-learning across all time periods. •Transfer support actions were least implemented after training compared to before and during training.
5 (Section 5.2)	•Assessing differences in objective and subjective training success between e-learning and classroom training. •Examining what type of knowledge–factual or applied–is best achieved in each training setting.	•Field experiment with time-lag design •Randomized •Corporate environment •Direct comparison between e-learning and classroom training, no control group •Same content learned parallel in e-learning and classroom training •Objective (performance test measures: factual and applied knowledge) and subjective training success measures •*N* = 86	•Overall performance test results did not vary across training settings but across type of knowledge. •Factual knowledge was more effectively trained in e-learning. •Applied knowledge was more effectively trained in classroom training only immediately after training, no difference occurred across training settings six to eight weeks later. •Success of learning and transfer did not vary across training settings. •High correlations between objective and subjective training success measures.
6 (Section 5.3)	•What is the key determinant of success of learning and transfer in corporate e-learning?	•Corporate environment •MASEM •12 samples with *N* = 974	•Transfer motivation as the most substantial key to ensure success of learning and transfer in corporate e-learning.

6.2 Practical implications for career construction across the life span

In the 21st century, career construction is a lifelong process. Due to the complex work environment that is characterized by constant changes, individuals' career pathways have become less stable and more unpredictable (e.g., Blustein, 2006; Savickas, 2004). To cope with unpredictable changes and to take advantage of arising opportunities, individuals need to actively engage in career construction processes. These processes involve that individuals need to make career-related choices in early and later career stages, which are among the most prevalent vocational problems. Inappropriate career-related behaviors incur financial (e.g., investing in training) as well as psychological costs (e.g., being dissatisfied with the job). As a consequence, many individuals face difficulties in making career-related choices (e.g., Amir, Gati, & Kleiman, 2008; Osipow, 1999). These difficulties can derive from a lack of knowledge or capability in dealing with one's own vocational interests and can lead to undesirable behaviors such as avoiding, halting, or taking suboptimal career decisions (Gati, Krausz, & Osipow, 1996). Career guidance and counseling aims at supporting clients making better career-related choices. According to this framework, different occupations are involved in guiding career construction processes of individuals across the life span. As different developmental tasks are addressed in different career stages, it seems useful to distinguish between guidance relevant in early career stages (i.e., career counselors) and in later career stages (i.e., T&D professionals).

In early career stages individuals need to take career choices and therefore strive for career direction. In this regard, career counselors can help individuals define their preferences by "transforming past experiences (successes and failures, satisfying and frustrating experiences) into specific preferences (or dislikes) for work-related activities and a self-understanding of one's skills, capacities, interests, and values (Van Esbroeck et al., 2005)" (Gati & Tal, 2008, p. 161). A wide variety of tools, often tests, assist career counselors' work to enable optimal and effective career choices. For example, assessments prior to entering university, usually so-called self-assessments, help individuals choose from the variety of occupations or academic majors in terms of narrowing down the variety of career options to the most suitable ones, based on their vocational interests (i.e., vocational interest measures) and abilities (i.e., ability measures). The presented research implies several implications for practitioners. First, as cognitive abilities and vocational interests were interrelated (Study 1), counselors should take both cognitive ability measures (such as GPA, ability-specific high school grades, results from college admission tests, or cognitive ability tests) and vocational interests into consideration. Results of such tests combined with further guidance from career counselors help individuals acquire decision-making skills in the long run. Taking both measures into account, this coincides with a current effort to establish vocational interest tests and cognitive ability tests across Germany. For example, completing an orientation test that comprises a vocational interest and a cognitive ability test (i.e., was-studiere-ich.de) is mandatory for all college applicants in Baden-Württemberg since the winter term 2011/12. Similarly, several German universities have implemented their own specific vocational interest tests, often in combination with cognitive ability tests. To reach the outlined goal, i.e. to support students in taking optimal career choices, such tests have to bear three important criteria in mind. Such tests need to be (a) grounded in a solid theoretical foundation, (b) both reliable and valid including aspects of construct, structural, content, and predictive validity, and (c) gender-fair. Especially for the third criterion, Studies 2 and 3 demonstrated that removing items showing large gender-specific DIF slightly reduced prediction bias (Study 2) and eliminating items showing gender-specific DIF had no considerable influence on the instrument’s psychometric structure (Study 3).

With regard to test fairness, large mean differences between men and women were consistently found for vocational interests across the globe (e.g., Su, Rounds, & Armstrong, 2009). Consequently, it is of utmost importance to correct for these biases to provide gender-fair vocational interest tests. Even though DIF methods have been widely used for bias correction in the ability domain, the presented studies are among the pioneers to apply this sophisticated method to vocational interest measures. Taking DIF analyses into account enhances the understanding of gender differences in vocational interests that are caused by measurement bias. Research on DIF in interest inventories indicates that there are other factors than latent interests that are responsible for gender differences found in vocational interests (Aros, Henly, & Curtis, 1998; Einarsdóttir & Rounds, 2009; W.-C. Wang, 2008). When DIF is detected, it indicates that response probabilities for an item differ in members of two groups (e.g., females and males), even though their underlying trait level is equal (P. W. Holland & Wainer, 1993; Osterlind & Everson, 2009). Thus, gender-specific DIF indicates a measurement bias which might also lead to impaired construct validity or different measurement properties for females and males. Results of Studies 2 and 3 demonstrated that DIF is one possibility to control for measurement bias and to significantly improve test fairness when developing interest inventories. Hence, test developers need to apply this method and councelors should be aware of potential shortcomings in test fairness and should look for high quality tests which take such issues into account.

In later career stages, individuals must adapt their career to respond to rapidly changing labor needs of post-modern societies. Hence, individuals focus on active career development for instance through participation in trainings.

Drawing the attention to training evaluation in this dissertation, numerous practical implications arise from the presented studies. Generally, T&D professionals should support individuals in their strive to be(come) successful in their occupation by actively guiding their learning process and maximizing their success of learning and success of transfer. In this respect, previous research in the field of T&D has identified a variety of actions for maximizing training effectiveness that are applicable to practice (e.g., Salas et al., 2012).

To further support the practical application of previous results, Study 4 provided an overview of the current prevalence of relevant practices and helps T&D professionals to identify development potential in their training settings. Here, similarities and dissimilarities between corporate e-learning and classroom training lead to the specific implementation of each training setting. More specifically, the results imply that more awareness and improvements are needed for implementing actions to maximize training effectiveness (a) in the phase after training in general (e.g., evaluating trainings), (b) for e-learning settings during all three phases, i.e. before, during, and after training, when compared to classroom training settings, and (c) that match and meet the needs of elderly employees. Moreover, the adequate implementation of actions that are especially efficient for e-learning settings helps to fully exploit the potential of web-based training beyond the (admittedly charming) use of simulations in e-learning in general. In order to facilitate the transfer from theory to practice, this dissertation has also taken a specific first step into the direction of putting theory into practice in that all participating companies were provided with a benchmark and a checklist of evidence-based recommendations and best practices for maximizing training effectiveness in the workplace. The hope is fulfilled that the results of the study with enclosed recommendations contribute to the practical impact and provide further practical insights for investigated companies to achieve the overarching goal of promoting training effectiveness and maximizing the success of transfer in the workplace. In this respect, the results can also be used by other companies to determine their relative standing in terms of facilitating training success and enable them to infer precise ways to further improve their trainings.

With respect to the questions whether classroom training or e-learning settings are preferable for trainings, no clear answer can be given. According to the results of Study 5, e-learning settings appear to be superior for learning factual knowledge, whereas classroom training settings should be chosen for applied knowledge. Combining both types of training settings, blended learning as a mixed training setting can be speculated to be the key to maximize training success in term of both factual and applied knowledge. Importantly, when choosing an e-learning setting, specific actions to support training success deem especially appropriate. In this regard, the meta-analysis of Study 6 identified important drivers of successful training transfer, thereby, also speaking to the highly relevant transfer problem. More specifically, the results highlight various leverages for HRD practice: First, before training, T&D professionals should raise trainees' awareness that effort devoted to transfer learning will lead to changes in job performance. Second, (virtual) trainers should incorporate specific motivational factors in the learning process during training and design transfer-supportive training. They could, for instance, make sure that learning content, examples, and instructions in the training environment match the transfer environment. Finally, supervisors should provide feedback after training to further enhance success of transfer.

As training success seems important to individuals' career development and T&D professionals have the power to actively influence the training and transfer environment, they should foster an active influence on trainee characteristics (i.e., motivation to transfer) and transfer design (according to Baldwin & Ford's [1988] categories). Therefore, variables that enhance success of learning and success of transfer need to be promoted. In this respect, the results of Study 6 can help T&D professionals to better guide their company's HRD in terms of facilitating and improving training transfer so that organizations' resources (e.g., time and money) are used efficiently and not wasted (Latham, 2007, p. 3). Taking these recommendations and leverage points into account will ultimately enhance the company's performance and yield a competitive advantage in the turbulent modern markets.

6.3 Strengths and limitations

Strengths

"Career-related choices are among the most important decisions people make during their lifetime" (Gati & Tal, 2008, p. 157). As these decisions affect a variety of individuals' aspects of life on a long-term basis (e.g., their lifestyle, well-being, economic and social status, sense of personal productivity, and contribution to society [Gati & Tal, 2008]), it is necessary for each individual to engage in well-founded career choices (e.g., Super, 1980). In early career stages, career choices primarily lead individuals' future career pathways, whereas in later career stages, career choices have smaller impacts on career construction (e.g., deciding for specific training activities).

From a theoretical point of view, the major strength of this dissertation lies in the combination of traditional and new theoretical approaches in vocational psychology by referring to the overarching career construction theory (Savickas, 2005).). From the perspective of rather traditional person-environment fit theories (e.g., Dawis & Lofquist, 1984; J. L. Holland, 1997), initial career choices are regarded as successful when a high congruence between individuals' interests and their environment exists. In contrast, more recent career development theories (e.g., L. S. Gottfredson, 2004; Super, 1990) suggest that career decisions depend on developmental changes which occur in individuals' preferences, career maturity, and adaptability. Combining these two fundamentally different outlooks on careers, the rationale of career construction theory illustrates how individuals actively construct their careers by using their vocational personality to adapt to the changing nature of work environment. Building on this general view of career construction theory, this dissertation explores key elements of career construction processes during the two most prominent developmental career stages, which are crucial for career success. Thereby, this dissertation moves beyond the focus on single phases or a specific career decision prevalent in most previous research and provides a more overarching view on lifelong career success instead.

From a content-related point of view, all studies contributed to current research issues: First, studies on vocational interests contributed to the continuous question of (a) how vocational interests and specific cognitive abilities are related and (b) how interest inventories can suspend gender differences in instrument validity and prediction without losing structural validity, which refers directly to the ongoing debate regarding test fairness (AERA, APA, & NCME, 1999; NIE, 1975). Second, studies on training evaluation contributed to ensure training success, thus speaking to the so-called transfer problem of trainings in the workplace which is one of the key issues in training research (e.g., Baldwin & Ford, 1988; Saks et al., 2014). In this respect, the dissertation takes a two-step approach. First, the prevalence of transfer support actions (derived from previous research) in corporate practice is assessed among several German companies. Second, key predictors of success of learning and success of transfer are identified with a specific emphasis on e-learning in order to increase empirical evidence and enhance their application in practice. In this respect, this issue had been raised in the previous training literature (Aguinis & Kraiger, 2009). For example, Study 5 experimentally contributed to training effectiveness by specifically comparing e-learning and classroom training settings. Similarly, Study 6 was designed to shed more light on transfer support actions effective (e.g., training design and motivation to transfer) in e-learning settings, which have previously been investigated almost exclusively in traditional classroom settings. Thus, this dissertation is in line with the current shift from traditional classroom training to e-learning settings and answers the current need to better understand the particularities of corporate e–learning. As this dissertation picked up training transfer as an important topic, it further closely relates to current investigations of The International Journal of Training and Development that published its first special issue on training transfer in 2014 (Saks et al., 2014).

A further strength of this dissertation is the wide variety of methods applied for the different studies: They span from confirmatory factor analysis and randomization tests of hypothesized order relations to classical meta-analysis and MASEM applied in a multi-sample study. The meta-analytical approach allows the researcher to statistically combine results from different studies to identify patterns among study results. This goes hand in hand with higher statistical power compared to measures derived from a single study and often with higher external validity as well, as several samples from different situations are studied to derive the results. In this respect, Study 1 presents a classical meta-analytic approach according to Hunter et al. (1982) to aggregate the rich knowledge on vocational interests and cognitive abilities established in previous studies. Study 6, on the contrary, uses the meta-analytic approach to combine results on predictors facilitating training success from several studies conducted within this dissertation and uses these results for further analyses through a meta-analytic structural equation model (MASEM, Geiser, 2010). A MMR framework was used in Study 2 to assess differential prediction of the specific dominant letter on person-environment fit. Confirmatory factor analyses and randomization tests of hypothesized order relations as the core analytic methods (Rounds et al., 1992) were used in Study 3. Descriptive analyses and inductive analyses (analyses of variance, *t*-tests) were used in Study 4 and Study 5.

Finally, this dissertation comprises a wide spectrum of investigated samples. In studies on vocational interests, primarily students were assessed, whereas in studies on training evaluation, the perspectives of T&D professionals (Study 4) and employees as trainees were explored (Study 5, Study 6) in diverse companies which varied in branches of industries. Taken as a whole, all investigated samples stem from real world settings.

Overall, this dissertation contributes to the continuous dialogue between theories that are important in career construction processes across the life span and practical needs of counselees to be assisted for a successful continuous career construction.

Limitations

As with all research, the studies of this dissertation entail several limitations, which might point to promising avenues for future research. In this respect, there are three important limitations that arose across studies. First, sufficient sample sizes are always important to guarantee both high statistical power and representativeness of the test. For the meta-analyses (Studies 1 and 6), the sample sizes of the meta-analysis presented in Study 1 were in line with similar previous research. Similarly, sample sizes in Study 6 were already impressive, given the practical difficulties to assess company data from real world e-learning training settings. The other two studies on vocational interests (Studies 2 and 3) demonstrated a sufficient sample size for the analyses carried out. However, enlarging the response rate (27.5%) in Study 4 (e.g., by providing other incentives), and increasing the number of trainees in the experimental study (Study 5) may lead to more generalizability and is associated with higher statistical power, thus, making the test of hypotheses less conservative (Kline, 2011). Taken together, although the presented sample sizes are acceptable for the research and were sufficient to detect important effects, larger samples sizes would be desirable to assure sufficient statistical power and enlarge the representativeness of the samples.

Second, most of the studies rely on self-reporting from students (Studies 1, 2, and 3), T&D professionals (Study 4), and trainees (Studies 5 and 6). Even though objective data is preferable, they are very hard to gather in the corporate environment. As objective and subjective measures strongly correlate (i.e., in terms of training success measures in Study 5), relying on subjective data seemed an adequate method. However, the results bear the risk that the obtained relationships among variables are inflated through common method bias (e.g., Podsakoff et al., 2003; Spector, 2006), which is a crucial weakness of Study 6. The common method bias is less likely in Study 5 because objective data was additionally assessed, but it can be excluded in Study 4 as various T&D professionals for each company were assessed. However, some constructs can hardly be assessed from the outside, so that self-ratings are a straightforward choice for constructs such as satisfaction or subjective person-environment fit (Study 2). With regard to this, perceptions and cognitions of workplace behaviors and internal states are best assessed through self-reports (e.g., Markóczy, 1997). To prevent this bias, gathering data with multiple methods should be investigated, for example, by involving not only one selected group of people (e.g., T&D professionals) but also by assessing the perspective of trainees with multiple sources, colleagues of trainees, supervisors, or customers. However, due to specific data-collection restrictions of the cooperating companies in Studies 5 and 6 (companies of each primary study), ratings of others (e.g., supervisors) than the trainees were not feasible, which could be an important addition in future research.

Third, this dissertation is limited in its cultural generalizability as the majority of investigated samples is German. A noteworthy exception is the meta-analysis presented in Study 1, which includes samples from Australia, Austria, Croatia, Germany, and the United States. Similarly, the companies involved in Study 4 and some of the companies involved in Study 6 had an international scope of training application, so that the data might not be limited exclusively to Germany. Therefore, even though most data collection was carried out in Germany, it is highly probable that the results can be generalized at least to culturally close Western countries. Future research should verify whether similar patterns can be found across the globe.

In sum, bigger samples with a larger variety of target groups that are not as limited to a German population and that are homogeneously distributed (e.g., in terms of gender) is suggested in further research investigations in order to cross-culturally generalize the results found. In spite of the discussed limitations, this dissertation makes important contributions to the ongoing debates in career choice and career development and can provide a sophisticated basis for future research, which should aim to overcome the specific shortcomings presented here.

6.4 Directions for future research

The empirical evidence presented in this dissertation contributes to our knowledge and understanding vocational interests and training transfer as important issues in the current career construction debate. Beyond the insights and implications which can be directly derived from the results, they can also spur further research efforts on both assessing vocational interests fairly to guide career choices and examining means to ensure training success especially in e-learning settings to support career development in later career stages. In this regard, the particular findings of the six studies and their inherent limitations point to a variety of promising avenues for future research.

First, as to vocational interests, all of the presented conclusions make reference to Holland’s hexagonal model as it is still the leading structural model dominating the field of vocational interests. Furthermore, Holland’s assumptions have been widely validated (e.g., for a meta-analysis, see T. J. Tracey & Rounds, 1993) and are generalizable across gender (Darcy & Tracey, 2007), and age (Darcy & Tracey, 2007). Nevertheless, empirical evidence of Holland’s structural conceptualization varies considerably across cultures (e.g., Long & Tracey, 2006; Rounds & Tracey, 1996). Hence, culturally driven DIF-reductions would be an important next step to make vocational interest tests applicable across cultures. Moreover, considering gender and cultural variables simultaneously would further increase our understanding of how structural properties of interest measures might be influenced by eliminating items showing large gender-specific DIF. Thus, the yielded results imply that similar corrections are needed for other Holland-based interest instruments (e.g, the Strong Interest Inventory by Harmon et al., 1994) and potentially for different conceptualizations of vocational interests (e.g., the hierarchical model by Gati’s (1991) in general.

Second, as to training evaluation, assessing objective and subjective training success longitudinally at several time points across a longer interval between training and criterion assessment would be an important next step to demonstrate stable transfer effects over a longer period of time. In that case, the assessment of more frequent applications of transfer activities can be facilitated which can help us to understand how training success develops over time, which ultimately also speaks to the question whether organizational goals were achieved through the particular training. Another issue refers to investigating a control group in experimental designs. Unfortunately, implementing real experimental designs in the corporate environment often remains wishful thinking due to companies' restrictions (e.g., companies' equality guidelines). However, using two training groups with the "control group" starting their training after the training of the experimental group has ended might be a possibility to strive for experimental setups. Finally, as outlined above, future research should work towards larger and more heterogeneous corporate samples, especially in corporate e-learning settings.

Third, taking the life span approach more serious, proposing and investigating a framework that comprises vocational interests, cognitive abilities, and additionally personality traits as well as outcomes of training across the life span would be highly recommended. In line with previous cross-sectional research (e.g., Ackerman & Heggestad, 1997; Anthoney & Armstrong, 2010; Armstrong et al., 2008; Armstrong & Rounds, 2010), the meta-anlysis in Study 1 demonstrated that the relationship between interests and specific cognitive abilities becomes more pronounced with age. However, as the meta-analysis relied exclusively on cross-sectional primary data, causal interferences cannot be drawn. Consequently, further studies should investigate the reciprocal relationship between vocational interests and crystallized intelligence (i.e., the acquisition of knowledge, skills, and aptitudes) depending on possible moderators such as specialization in education. In addition, the assessment of intra-individual developmental processes based on the changes of individual vocational interests and cognitive ability profiles enables a true developmental perspective across the life span. In this respect, previous studies have revealed that knowledge, skills, and vocational interests may change over time, whereas general mental abilities and personality traits are stable throughout adulthood (Kanfer & Ackerman, 2004). More specifically, J. L. Swanson (1999) showed that vocational interests are particularly stable only after the age of 30. This indicates that differences in vocational interest profiles rather stem from being exposed to different environments than from changes in interest patterns over time. With regard to individuals' adaptability to changes at work which further leads to organizational adaptability (Ployhart & Turner, 2014), a large overarching model including developmental changes in individuals' interests and their impact on career success seems to open a rich avenue for future research. A truly overarching model should reciprocally consider vocational interests and training outcomes and allow for smoother transitions across stages that are not limited to a specific age range. As stages are not necessarily sequential in post-modern societies, individuals engage in further training not only in later but also in early career stages, and, in turn, individuals engage in a “new” career direction in later career stages. Therefore, an overall model should be adaptive and thus contribute to individuals’ adaptability which is one of the key factors for individuals’ career success (Chan, 2014, p. XXI). More than ever before, it is necessary to adapt quickly to the changing nature of work environments in the 21st century which helps to successfully construct careers across the life span.

References

References marked with an asterisk indicate studies included in the meta-analysis (Study 1).

Achter, J. A., Lubinski, D., Benbow, C. P., & Eftekhari-Sanjani, H. (1999). Assessing vocational preferences among gifted adolescents adds incremental validity to abilities: A discriminant analysis of educational outcomes over a 10-year interval. *Journal of Educational Psychology, 91*(4)*,* 777-786. doi:10.1037/0022-0663.91.4.777

Ackerman, P. L. (1996). A theory of adult intellectual development: Process, personality, interests, and knowledge. *Intelligence*, *22*(2), 227-257. doi:10.1016/S0160-2896(96)90016-1

Ackerman, P. L. (1997). Personality, self-concept, interests, and intelligence: Which construct doesn't fit? *Journal of Personality, 65*(2), 171-204. doi:10.1111/j.1467-6494.1997.tb00952.x

*Ackerman, P. L. (2000). Domain-specific knowledge as the 'dark matter' of adult intelligence: Gf/Gc, personality and interest correlates. *The Journals of Gerontology: Series B: Psychological Sciences and Social Sciences, 55*(2), 69-84. doi:10.1093/geronb/55.2.P69

Ackerman, P. L., & Beier, M. E. (2003). Intelligence, personality, and interests in the career choice process. *Journal of Career Assessment*, *11*(2), 205-218. doi:10.1177/1069072703011002006

*Ackerman, P. L., Bowen, K. R., Beier, M., & Kanfer, R. (2001). Determinants of individual differences and gender differences in knowledge. *Journal of Educational Psychology, 93*(4), 797-825. doi:10.1037//0022-0663.93.4.797

Ackerman, P. L., & Heggestad, E. D. (1997). Intelligence, personality, and interests: Evidence for overlapping traits. *Psychological Bulletin, 121*(2)*,* 219-245. doi:10.1037/0033-2909.121.2.219

*Ackerman, P. L., Kanfer, R., & Goff, M. (1995). Cognitive and noncognitive determinants and consequences of complex skill acquisition. *Journal of Experimental Psychology: Applied, 1*(4)*,* 270-304. doi:10.1037/1076-898X.1.4.270

Aguinis, H. (2004). *Regression analysis for categorial moderator*s. New York, NY: Guilford Press.

Aguinis, H., Culpepper, S. A., & Pierce, C. A. (2010). Revival of test bias research in employment testing. *Journal of Applied Psychology*, *95*(4), 648-680. doi:10.1037/a0018714

Aguinis, H., & Kraiger, K. (2009). Benefits of training and development for individuals and teams, organizations, and society. *Annual Review of Psychology*, *60*(1), 451-474. doi:10.1146/annurev.psych.60.110707.163505

Aguinis, H., & Stone-Romero, E. F. (1997). Methodological artifacts in moderated multiple regression and their effects on statistical power. *Journal of Applied Psychology*, *82*(1), 192-206. doi:10.1037//0021-9010.82.1.192

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Thousand Oaks, CA: Sage.

Alfonso, V. C., Flanagan, D. P., & Radwan, S. (2005). The impact of the Cattell-Horn-Carroll theory on test development and interpretation of cognitive and academic abilities. In D. Flanagan & P. Harrison (Eds.), *Contemporary intellectual assessment, theories, tests, and issues* (pp. 185-202). New York, NY: Guilford Press.

Alliger, G. M., & Janak, E. A. (1989). Kirkpatrick's levels of training criteria: Thirty years later. *Personnel Psychology*, *42*(2), 331-342. doi:10.1111/j.1744-6570.1989.tb00661.x

Alliger, G. M., Tannenbaum, S. I., Bennett, W., Traver, H., & Shotland, A. (1997). A meta-analysis of the relations among training criteria. *Personnel Psychology*, *50*(2), 341-358. doi:10.1111/j.1744-6570.1997.tb00911.x

American College Testing (ACT) Program. (1995). *Technical manual: Revised Unisex Edition of the ACT interest inventory (UNIACT-R).* Iowa City, IA: ACT Inc.

American College Testing (ACT). (2009). *Interest inventory technical manual*. Iowa City, IA: Author.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (AERA, APA, & NCME). (1999). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.

American Society for Training and Development (ASTD). (2012). *The 2012 state of the industry report*. Retrieved from http://store.astd.org/Default.aspx?tabid=%20167& ProductId=23805

Amir, T., Gati, I., & Kleiman, T. (2008). Understanding and interpreting career decision-making difficulties. *Journal of Career Assessment*, *16*(3), 281-309. doi:10.1177/1069072708317367

Anderson, M. Z., Tracey, T. J., & Rounds, J. (1997). Examining the invariance of Holland's vocational interest model across gender. *Journal of Vocational Behavior*, *50*(3), 349-364. doi:10.1006/jvbe.1996.1550

Armstrong, P. I., Hubert, L., & Rounds, J. (2003). Circular unidimensional scaling: A new look at group differences in interest structure. *Journal of Counseling Psychology*, *50*(3), 297-308. doi:10.1037/0022-0167.50.3.297

Armstrong, P. I., Su, R., & Rounds, J. (2011). Vocational interests: The road less traveled. In T. Chamorro-Premuzic, S. Von Stumm, & A. Furnham (Eds.), *Wiley-Blackwell handbooks in personality and individual differences. The Wiley-Blackwell handbook of individual differences* (pp. 608-631). Malden, MA: Wiley-Blackwell.

Aros, J. R., Henly, G. A., & Curtis, N. T. (1998). Occupational sextype and sex differences in vocational preference-measured interest relationships. *Journal of Vocational Behavior*, *53*(2), 227-242. doi:10.1006/jvbe.1997.1612

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. doi:10.1037/0021-9010.88.2.234

Assouline, M., & Meir, E. I. (1987). Meta-analysis of the relationship between congruence and well-being measures. *Journal of Vocational Behavior*, *31*(3), 319-332. doi:10.1016/0001-8791(87)90046-7

Awoniyi, E. A., Griego, O. V., & Morgan, G. A. (2002). Person-environment fit and transfer of training. *International Journal of Training and Development*, *6*(1), 25-35. doi:10.1111/1468-2419.00147

Baldwin, T. T., & Ford, J. K. (1988). Transfer of training: A review and directions for future research. *Personnel Psychology*, *41*(1), 63-105. doi:10.1111/j.1744--6570.1988.tb00632.x

Baldwin, T. T., Ford, J. K., & Blume, B. D. (2009). Transfer of training 1988-2008: An updated review and agenda for future research. In P. G. Hodgkinson & J. K. Ford (Eds.), *International review of industrial and organizational psychology* (Vol. 24, pp. 41-70). Chichester, UK: Wiley. doi:10.1002/9780470745267.ch2

Baldwin, T. T., & Magjuka, R. J. (1991). Organizational training and signals of importance: Linking pretraining perceptions to intentions to transfer. *Human Resource Development Quarterly*, *2*(1), 25-36. doi:10.1002/hrdq.3920020106

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*(2), 191-215. doi:10.1037//0033-295X.84.2.191

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory.* Englewood Cliffs, NJ: Prentice-Hall.

Bandura, A. (Ed.). (1995). *Self-efficacy in changing societies*. Cambridge, UK: Cambridge University Press.

Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, *44*(1), 1-26. doi:10.1111/j.1744-6570.1991.tb00688.x

Barrick, M. R., & Mount, M. K. (2005). Yes, personality matters: Moving on to more important matters. *Human Performance, 18*(4), 359-372. doi:10.1207/s15327043hup1804_3

Barrick, M. R., Mount, M. K., & Gupta, R. (2003). Meta-analysis of the relationship between the five-factor model of personality and Holland's occupational types. *Personnel Psychology, 56*(1), 45-74. doi:10.1111/j.1744-6570.2003.tb00143.x

Bateman, T. S., & Crant, J. M. (1993). The proactive component of organizational behavior: A measure and correlates. *Journal of Organizational Behavior*, *14*(2), 103-118. doi:10.1002/job.4030140202

Bates, R. A., Holton, E. F., Seyler, D. L., & Carvalho, M. A. (2000). The role of interpersonal factors in the application of computer-based training in an industrial setting. *Human Resource Development International*, *3*(1), 19-42. doi:10.1080/136788600361920

Bates, R. A., Kauffeld, S., & Holton, E. F. (2007). Examining the factor structure and predictive ability of the German-version of the Learning Transfer Systems Inventory (LTSI). *Journal of European Industrial Training, 31*(3), 195-211. doi:10.1108/03090590710739278

Bates, R. A., & Khasawneh, S. (2005). Organizational learning culture, learning transfer climate and perceived innovation in Jordanian organizations. *International Journal of Training and Development*, *9*(2), 96-109. doi:10.1111/j.1468-2419.2005.00224.x

Baumeister, R. F., & Finkel, E. J. (Eds.). (2010). *Advanced social psychology: The state of the science*. New York, NY: Oxford University Press.

Bausch, S., Sonntag, K., Stegmaier, R., & Noefer, K. (2010). Können Ältere mit neuen Medien lernen? Gestaltung und Evaluation eines e-Learning Behavior Modeling Trainings für verschiedene Altersgruppen [Do older learners benefit from new media? Design and evaluation of an e-learning behavior modeling training for different age groups]. *Zeitschrift für Arbeitswissenschaft*, *64*(3), 239-251.

Bedwell, W. L., & Salas, E. (2010). Computer-based training: Capitalizing on lessons learned. *International Journal of Training and Development*, *14*(3), 239-249. doi:10.1111/j.1468-2419.2010.00355.x

Beinicke, A., Pässler, K., & Hell, B. (2014). Does gender-specific differential item functioning affect the structure in vocational interest inventories? *International Journal for Educational and Vocational Guidance*, *14*(2), 181-198. doi:10.1007/s10775-013-9254-y

Bell, B. S., & Kozlowski, W. J. (2002). Goal orientation and ability: Interactive effects on self-efficacy, performance, and knowledge. *Journal of Applied Psychology*, *87*(3), 497-505. doi:10.1037/0021-9010.87.3.497

Berenbaum, S. A., Baxter, L., Seidenberg, M., & Hermann, B. (1997). Role of the hippocampus in sex differences in verbal memory: Memory outcome following left anterior temporal lobectomy. *Neuropsychology*, *11*, 585-591. doi:10.1037/0894-4105.11.4.585

Berge, Z. L., & Giles, L. (2008). Implementing and sustaining e-learning in the workplace. *International Journal of Web-Based Learning and Teaching Technologies*, *3*(3), 44-53. doi:10.4018/jwltt.2008070104

Bergmann, C., & Eder, F. (2005). *Allgemeiner Interessen-Struktur-Test mit Umwelt-Struktur-Test (AIST-R/UST-R)–Revision.* [General Interest Structure Test and Environmental Structure Test–Revision]. Göttingen, Germany: Beltz Test.

Best Workplaces in Health Services. (2012, February 3). Retrieved from http://www.presseportal.de/pm/69829/2192230/beste-arbeitgeber-im-gesundheits wesen-2012-ausgezeichnet

Biernat, M., & Sesko, A. K. (2013). Evaluating the contributions of members of mixed-sex work teams: Race and gender matter. *Journal of Experimental Social Psychology*, *49*(3), 471-476. doi:10.1016/j.jesp.2013.01.008

Blume, B. D., Ford, J. K., Baldwin, T. T., & Huang, J. L. (2010). Transfer of training: A meta-analytic review. *Journal of Management*, *36*(4), 1065-1105. doi:10.1177/0149206309352880

Blustein, D. L. (2006). *The psychology of working: A new perspective for career development, counseling and public policy*. doi:10.4324/9780203935477

Bonk, C. J., Kim, K.-J., & Zeng T. (2005). Future directions of blended learning in higher education and workplace learning settings. In C. J. Bonk & C. R. Graham (Eds.), *Handbook of blended learning: Global perspectives, local designs* (pp. 550-568). San Francisco, CA: Pfeiffer.

Bortz, J., & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* [Research methods and evaluation for social scientists]. Berlin, Germany: Springer.

Brown, S. D., & Gore, P. A. (1994). An evaluation of interest congruence indices: Distribution characteristics and measurement properties. *Journal of Vocational Behavior*, *45*, 310-327. doi:10.1006/jvbe.1994.1038

Brown, S. D., & Lent, R. W. (Eds.). (2004). *Career development and counseling: Putting theory and research to work*. Hoboken, NJ: Wiley.

Browne, M. W. (1992). Circumplex models for correlation matrices. *Psychometrika*, *57*(4), 469-497. doi:10.1007/BF02294416

Browne, M. W. (1995). *CIRCUM: Satellite program to AUFIT for circumplex models*. Columbus, OH: Psychology Department, Ohio State University.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage.

Bubany, S. T., & Hansen, J.-I. C. (2011). Birth cohort change in the vocational interests of female and male college students. *Journal of Vocational Behavior, 78*(1)*,* 59-67. doi:10.1016/j.jvb.2010.08.002

Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion* [Introduction into test construction and questionnaires]. München, Germany: Pearson.

Bürg, O., & Mandl, H. (2005). Akzeptanz von E-Learning in Unternehmen [Acceptance of e-learning in companies]. *Zeitschrift für Personalpsychologie*, *4*(2), 75-85. doi:10.1026/1617-6391.4.2.75

Burke, L. A., & Hutchins, H. M. (2007). Training transfer: An integrative literature review. *Human Resource Development Review*, *6*(3), 263-296. doi:10.1177/1534484307303035

Burke, L. A., Hutchins, H. M., & Saks, A. (2013). Best practices in training transfer. In M. Paludi (Ed.), *The psychology for business success* (Vol. 3, pp. 115-132). Westport, CT: Praeger.

Cannon, M. D., & Witherspoon, R. (2005). Actionable feedback: Unlocking the power of learning and development. *Academy of Management Executive*, *19*(2), 120-134. doi:10.5465/ame.2005.16965107

*Carless, S. A. (1999). Career assessment: Holland's vocational interests, personality characteristics, and abilities. *Journal of Career Assessment, 7*(2), 125-144. doi:10.1177/106907279900700203

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge, UK: Cambridge University Press.

*Carson, A. D. (1996). Aptitudes across Holland's types: Implications for school-based counsellors. *McGill Journal of Education, 31,* 319-332.

*Carson, A. D. (1998a). The integration of interests, aptitudes, and personality traits: A test of Lowman's matrix. *Journal of Career Assessment, 6*(1), 83-105. doi:10.1177/106907279800600106

*Carson, A. D. (1998b). The relation of self-reported abilities to aptitude test scores: A replication and extension. *Journal of Vocational Behavior, 53*(3), 353-371. doi:10.1006/jvbe.1998.1641

Cattell, R. B. (1971). *Abilities: Their structure, growth, and action*. Boston, MA: Houghton Mifflin.

Cattell, R. B. (1987). *Intelligence: Its structure, growth, and action* (Vol. 35). Amsterdam, Netherlands: Elsevier.

Caudron, S. (1999). Free agent learner. *Training and Development*, *53*(8), 27-31.

Cavanaugh, C. S. (2001). The effectiveness of interactive distance education technologies in K-12 learning: A meta-analysis. *International Journal of Educational Telecommunications*, *7*(1), 73-88.

Chan, D. (Ed.). (2014). *Individual adaptability to changes at work: New directions in research*. New York, NY: Routledge.

Chen, H. C., Holton, E. F., & Bates, R. A. (2005). Development and validation of the Learning Transfer System Inventory in Taiwan. *Human Resource Development Quarterly*, *16*(1), 55-84. doi:10.1002/hrdq.1124

Cheung, M. W. L., & Chan, W. (2005). Meta-analytic structural equation modeling: A two-stage approach. *Psychological Methods*, *10*, 40-64. doi:10.1037/1082-989X.10.1.40

Chiaburu, D. S., & Lindsay, D. R. (2008). Can do or will do? The importance of self-efficacy and instrumentality for training transfer. *Human Resource Development International*, *11*(2), 199-206. doi:10.1080/13678860801933004

Chiaburu, D. S., & Marinova, S. V. (2005). What predicts skill transfer? An exploratory study of goal orientation, training self-efficacy and organizational supports. *International Journal of Training and Development*, *9*(2), 110-123. doi:10.1111/j.1468-2419.2005.00225.x

Clarke, N. (2002). Job/work environment factors influencing training transfer within a human service agency: Some indicative support for Baldwin and Ford’s transfer climate construct. *International Journal of Training and Development*, *6*(3), 146-162. doi:10.1111/1468-2419.00156

Cleary, T. A. (1968). Test bias: Prediction of grades of negro and white students in integrated colleges. *Journal of Educational Measurement*, *5*, 115-124. doi:10.1111/j.1745-3984.1968.tb00613.x

Cleveland, J. N., & Shore, L. M. (1992). Self- and supervisory perspectives on age and work attitudes and performance. *Journal of Applied Psychology*, *77*(4), 469-484. doi:10.1037/0021-9010.77.4.469

Cohen, J. (1988). *Statistical power analysis for the behavioral science* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1)*,* 155-159. doi:10.1037/0033-2909.112.1.155

Cohen-Bendahan, C. C., Van de Beek, C., & Berenbaum, S. A. (2005). Prenatal sex hormone effects on child and adult sex-typed behavior: Methods and findings. *Neuroscience & Biobehavioral Reviews*, *29*(2), 353-384. doi:10.1016/j.neubiorev.2004.11.004

Collins, J. S., & Singh, V. (2006). Exploring gendered leadership. In D. McTavish & K. Miller (Eds.), *Women in leadership and management* (pp. 11-31). Cheltenham, UK: Elgar.

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. doi:10.1037/0021-9010.85.5.678

Cook, E. P., Heppner, M. J., & O'Brien, K. M. (2002). Career development of women of color and white women: Assumptions, conceptualization, and interventions from an ecological perspective. *The Career Development Quarterly*, *50(4)*, 291-305. doi:10.1002/j.2161-0045.2002.tb00574.x

Crant, J. M. (2000). Proactive behavior in organizations. *Journal of Management*, *26*(3), 435-462. doi:10.1177/014920630002600304

Cromwell, S. E., & Kolb, J. A. (2004). An examination of work-environment support factors affecting transfer of supervisory skills training to the workplace. *Human Resource Development Quarterly*, *15*(4), 449-471. doi:10.1002/hrdq.1115

Darcy, M. U. A. (2005). Examination of the structure of Irish students' vocational interests and competence perceptions. *Journal of Vocational Behavior*, *67*(2), 321-333. doi:10.1016/j.jvb.2004.08.007

Darcy, M. U. A., & Tracey, T. J. (2007). Circumplex structure of Holland's RIASEC interests across gender and time. *Journal of Counseling Psychology*, *54*(1), 17-31. doi:10.1037/0022-0167.54.1.17

Dawis, R. V. (1992). The individual differences tradition in counseling psychology. *Journal of Counseling Psychology, 39*(1), 7-19. doi:10.1037//0022-0167.39.1.7

Dawis, R. V. (2002). Person-environment-correspondence theory. In D. Brown & Associates (Eds.), *Career choice and development* (4th ed., pp. 427-464). San Francisco, CA: Jossey-Bass.

Dawis, R. V. (2005). The Minnesota theory of work adjustment. In S. D. Brown & R. W. Lent (Eds.), *Career development and counseling: Putting theory and research to work* (pp. 3-23). Hoboken, NJ: Wiley.

Dawis, R. V., & Lofquist, L. H. (1984). *A psychological theory of work adjustment* [data file]. Retrieved from http://www.appliedpsyj.org/paper/other/sfwang/1984A%20psychological%20theory%20of%20work%20adjustment%20%20%20An%20individual-differences%20model%20and%20its%20applications.pdf

Day, S. X, Rounds, J., & Swaney, K. (1998). The structure of vocational interests for diverse racial-ethnic groups. *Psychological Science*, *9*(1), 40-44. doi:10.1111/1467-9280.00007

Deary, I. J. (2003). Population sex differences in IQ at age 11: The Scottish mental survey 1932. *Intelligence, 31*(6)*,* 533-542. doi:10.1016/S0160-2896(03)00053-9

Deary, I. J., Egan, V., Gibson, G. J., Austin, E. J., Brand, C. R., & Kellaghan, T. (1996). Intelligence and the differentiation hypothesis. *Intelligence, 23*(2)*,* 105-132. doi:10.1016/S0160-2896(96)90008-2

Deary, I. J., Pattie, A., & Starr, J. M. (2013). The stability of intelligence from age 11 to age 90 years: The Lothian birth cohort of 1921. *Psychological Science, 24*(12)*,* 2361-2368. doi:10.1177/0956797613486487

Devos, C., Dumay, X., Bonami, M., Bates, R. A., & Holton, E. F. (2007). The Learning Transfer System Inventory (LTSI) translated into French: Internal structure and predictive validity. *International Journal of Training and Development*, *11*(3), 181-199. doi:10.1111/j.1468-2419.2007.00280.x

Dik, B. J., Hu, R. S. C., & Hansen, J.-I. C. (2007). An empirical test of the modified C index and SII, O*NET, and DHOC occupational code classifications. *Journal of Career Assessment*, *15*, 279-300. doi:10.1177/1069072707301203

Dittler, U. (2011). *E-Learning: Einsatzkonzepte und Erfolgsfaktoren des Lernens mit interaktiven Medien* [E-learning: Success criteria and implementation concepts with interactive media] (3rd ed.). doi:10.1524/9783486714432

Dringus, L. P. (2000). Towards active online learning: A dramatic shift in perspectives for learners. *The Internet and Higher Education*, *2*(4), 189-195. doi:10.1016/S1096-7516(00)00023-3

Dura, M., & Solga, M. (2010, September). *Proaktive Persönlichkeit und Trainingstransfer* [Proactive personality and transfer of training]. Presentation at the 47th congress of the German Psychological Society (DGPs), Bremen, Germany.

Eid, M., Gollwitzer, M., & Schmitt, M. (2011). *Statistik und Forschungsmethoden* [Statistics and research methods] (2nd ed.). Weinheim, Germany: Beltz.

Einarsdóttir, S., & Rounds, J. (2009). Gender bias and construct validity in vocational interest measurement: Differential item functioning in the Strong Interest Inventory. *Journal of Vocational Behavior*, *74*(3), 295-307. doi:10.1016/j.jvb.2009.01.003

Fabrigar, L. R., Visser, P. S., & Browne, M. W. (1997). Conceptual and methodological issues in testing the circumplex structure of data in personality and social psychology.

Personality and Social Psychology Review, *1*(3), 184-203. doi:10.1207/s15327957 pspr0103_1

Facteau, J. D., Dobbins, G. H., Russell, J. E. A., Ladd, R. T., & Kudisch, J. D. (1995). The influence of general perceptions for the training environment on pretraining motivation and perceived training transfer. *Journal of Management*, *21*(1), 1-25. doi:10.1016/0149-2063(95)90031-4

Fahrmeir, L., Künstler, R., Pigeot, I., & Tutz, G. (2011). *Statistik: Der Weg zur Datenanalyse* [Statistics: The way to data analysis]. Berlin, Germany: Springer.

Fisher, S. L., Wasserman, M. E., & Orvis, K. A. (2010). Trainee reactions to learner control: An important link in the e-learning equation. *International Journal of Training and Development*, *14*(3), 198-208. doi:10.1111/j.1468-2419.2010.00352.x

Flanagan, D. P., Alfonso, V. C., & Ortiz, S. O. (2012). The cross-battery assessment (XBA) approach: An overview, historical perspective, and current directions. In D. Flanagan & P. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 459-483). New York, NY: Guilford Press.

Flanagan, D. P., & Dixon, S. G. (2013). The Cattell-Horn-Carroll theory of cognitive abilities. In *Encyclopedia of Special Education* (pp. 368-382). Hoboken, NJ: Wiley.

Ford, J. K., Smith, E. M., Weissbein, D. A., Gully, S. M., & Salas, E. (1998). Relationships of goal orientation, metacognitive activity, and practice strategies with learning outcomes and transfer. *Journal of Applied Psychology*, *83*(2), 218-233. doi:10.1037/0021-9010.83.2.218

Ford, J. K. & Weissbein, D. A. (1997). Transfer of Training: An updated review and analysis. *Performance Improvement Quarterly*, *10*(2), 22-41. doi:10.1111/j.1937-8327.1997.tb00047.x

Fouad, N. A., & Dancer, L. S. (1992). Cross-cultural structure of vocational interests*. Journal of Vocational Behavior*, *40*(2), 129-143. doi:10.1016/0001-8791(92)90061-4

Frese, M., Fay, D., Hilburger, T., Leng, K., & Tag, A. (1997). The concept of personal initiative: Operationalization, reliability and validity in two German samples. *Journal of Occupational and Organizational Psychology*, *70*(2), 139-161. doi:10.1111/j.2044-8325.1997.tb00639.x

*Fritzsche, B. A., Powell, A. B., & Hoffman, R. (1999). Person-environment congruence as a predictor of customer service performance. *Journal of Vocational Behavior, 54*(1), 59-70. doi:10.1006/jvbe.1998.1645

Fry, K. (2001). E-learning markets and providers: Some issues and prospects. *Education and Training*, *43*(4/5), 233-239. doi:10.1108/EUM0000000005484

FTSE International Limited. (2014). *The Industry Classification Benchmark (ICB): Industry Structure and definitions* [data file]. Retrieved from http://www.icbenchmark.com/ICBDocs/Structure_Defs_English.pdf

Gallupe, B. (2001). Knowledge management systems: Surveying the landscape. *International Journal of Management Reviews*, *3*(1), 61-77. doi:10.1111/1468-2370.00054

Garrison, D. R., & Kanuka, H. (2004). Blended learning: Uncovering its transformative potential in higher education. *The Internet and Higher Education*, *7*(2), 95-105. doi:10.1016/j.iheduc.2004.02.001

Gati, I. (1979). A hierarchical model for the structure of vocational interests. *Journal of Vocational Behavior*, *15*(1), 90-106. doi:10.1016/0001-8791(79)90021-6

Gati, I. (1982). Testing models for the structure of vocational interests. *Journal of Vocational Behavior*, *21*(2), 164-182. doi:10.1016/0001-8791(82)90027-6

Gati, I. (1991). The structure of vocational interests. *Psychological Bulletin*, *109*(2), 309-324. doi:10.1037/0033-2909.109.2.309

Gati, I. (1998). Using career-related aspects to elicit preferences and characterize occupations for a better person-environment fit. *Journal of Vocational Behavior*, *52*(3), 343-356. doi:10.1006/jvbe.1997.1623

Gati, I., Krausz, M., & Osipow, S. H. (1996). A taxonomy of difficulties in career decision making. *Journal of Counseling Psychology*, *43*(4), 510-526. doi:10.1037/0022-0167.43.4.510

Gati, I., & Tal, S. (2008). Decision-making models and career guidance. In J. Athanasou & R. Van Esbroeck (Eds.), *International handbook of career guidance* (pp. 157-185). Berlin, Germany: Springer.

Gegenfurtner, A., Veermans, K., Festner, D., & Gruber, H. (2009). Integrative literature review: Motivation to transfer training: An integrative literature review. *Human Resource Development Review*, *8*(3), 403-423. doi:10.1177/1534484309335970

Geiser, C. (2010). *Datenanalyse mit Mplus: Eine anwendungsorientierte Einführung* [Data analysis with Mplus: An application-oriented introduction]. Wiesbaden, Germany: VS Verlag.

Germany's Best Workplaces 2012. (2012, March 9). Retrieved from http://www.presseportal.de/pm/69829 /2213029/deutschlands-beste-arbeitgeber-2012-ausgezeichnet

Gilpin-Jackson, Y., & Bushe, G. R. (2007). Leadership development training transfer: A case study of post-training determinants. *Journal of Management Development*, *26*(10), 980-1004. doi:10.1108/02621710710833423

Gist, M. E., Rosen, B., & Schwoerer, C. (1988). The influence of training method and trainee age on the acquisition of computer skills. *Personnel Psychology*, *41*(2), 255-265. doi:10.1111/j.1744-6570.1988.tb02384.x

Gnefkow, T. (2008*). Lerntransfer in der betrieblichen Weiterbildung: Determinanten für den Erfolg externer betrieblicher Weiterbildungen im Lern- und Funktionsfeld aus Teilnehmerperspektive [Transfer of learning in vocational training: Determinants for the success of external vocational training in the learning and functional area from participants' perspective]* (doctoral dissertation). Retrieved from http://pub.uni-bielefeld.de/luur/download?func=downloadFile&recordOId=2303980&fileOId=2303983

Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: On land and underwater. *British Journal of Psychology*, *66*(3), 325-331. doi:10.1111/j.2044-8295.1975.tb01468.x

Goldstein, I. L., & Ford, J. K. (2002). *Training in organizations: Needs assessment, development, and evaluation* (4th ed.). Belmont, CA: Wadsworth.

Gottfredson, G. D., & Holland, J. L. (1975). Vocational choices of men and women: A comparison of predictors from the self-directed search. *Journal of Counseling Psychology*, *22*, 28-34. doi:10.1037/h0076150

Gottfredson, G. D., & Holland, J. L. (1978). Toward beneficial resolution of the interest inventory controversy. In C. K. Tittle & D. G. Zytowski (Eds.), *Sex-fair interest measurement: Research and implications* (pp. 43-51). Washington, DC: U.S. Government Printing Office.

Gottfredson, L. S. (1981). Circumscription and compromise: A developmental theory of occupational aspirations. *Journal of Counseling Psychology*, *28*(6), 545-579. doi:10.1037/0022-0167.28.6.545

Gottfredson, L. S. (1986). Occupational aptitude patterns map: Development and implications for a theory of job aptitude requirements. *Journal of Vocational Behavior, 29*(2)*,* 254-291. doi:10.1016/0001-8791(86)90008-4

Gottfredson, L. S. (1997). Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography. *Special Issue Intelligence and Social Policy, 24*(1)*,* 13-23. doi:10.1016/S0160-2896(97)90011-8

Gottfredson, L. S. (2004). Applying Gottfredson's theory of circumscription and compromise in career guidance and counseling. In S. D. Brown & R. W. Lent (Eds.), *Career development and counseling: Putting theory and research to work* (pp. 71-100). Hoboken, NJ: Wiley.

Goy, R. W., & McEwen, B. S. (2007). Sexually dimorphic behavior: Definition and the organizational hypothesis. In G. Einstein (Ed.), *Sex and the brain* (pp. 7-13). Cambridge, MA: MIT Press.

Gravill, J., & Compeau, D. (2008). Self-regulated learning strategies and software training. *Information and Management*, *45*(5), 288-296. doi:10.1016/j.im.2008.03.001

Great Place to Work. (2014a, March 1). Retrieved from http://www.greatplacetowork.de/ueber-uns/ueber-uns-auf-einen-blick

Great Place to Work. (2014b). *The Great Place to Work model* [data file]. Retrieved from http://www.greatplacetowork.de/storage/documents/GPTW_Modell.pdf

Grohmann, A., Beller, J., & Kauffeld, S. (2014). Exploring the critical role of motivation to transfer in the training transfer process. *International Journal of Training and Development*, *18*(2), 84-103. doi:10.1111/ijtd.12030

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. doi:10.1111/ijtd.12005

Grossman, R., & Salas, E. (2011). The transfer of training: What really matters. *International Journal of Training and Development*, *15*(2), 103-120. doi:10.1111/j.1468-2419.2011.00373.x

Gupta, S., Tracey, T. J., & Gore, P. A., Jr. (2008). Structural examination of RIASEC scales in high school students: Variation across ethnicity and method. *Journal of Vocational Behavior*, *72*(1), 1-13. doi:10.1016/j.jvb.2007.10.013

Halpern, D. F. (2000). *Sex differences in cognitive abilities* (3rd ed.). Mahwah, NJ: Erlbaum.

Hammond, D. (2001) Reality bytes. *People Management*, *7*(2), 26-31.

Han, H.-J., Dick, G., Case, T., & Van Slyke, C. (2009). Key capabilities, components, and evolutionary trends in corporate e-learning systems. In H. Song & T. T. Kidd (Eds.), *Handbook of research on human performance and instructional technology* (pp. 446-469). doi:10.4018/978-1-60566-782-9.ch027

Hansen, J.-I. C., Collins, R. C., Swanson, J. L., & Fouad, N. A. (1993). Gender differences in the structure of interests. *Journal of Vocational Behavior*, *42*(2), 200-211. doi:10.1006/jvbe.1993.1014

Hansen, J.-I. C., & Dik, B. J. (2005). Evidence of 12-year predictive and concurrent validity for SII occupational scale scores. *Journal of Vocational Behavior*, *67*(3), 365-378. doi:10.1016/j.jvb.2004.08.001

Hansen, J.-I. C., & Neuman, J. L. (1999). Evidence of concurrent prediction of the Campbell Interest and Skill Survey (CISS) for college major selection. *Journal of Career Assessment, 7*(3)*,* 239-247. doi:10.1177/106907279900700304

Hansen, J.-I. C., & Swanson, J. L. (1983). Stability of interests and the predictive and concurrent validity of the 1981 Strong-Campbell Interest Inventory for college majors. *Journal of Counseling Psychology*, *30*(2), 194-201. doi:10.1037/0022-0167.30.2.194

Hansen, J.-I. C., & Tan, R. N. (1992). Concurrent validity of the 1985 Strong Interest Inventory for college major selection. *Measurement and Evaluation in Counseling and Development*, *25*(2), 53-57.

Hansford, B. C., & Hattie, J. A. (1982). The relationship between self and achievement/performance measures. *Review of Educational Research, 52*(1), 123-142. doi:10.3102/00346543052001123

Harmon, L. W., Hansen, J.-I. C., Borgen, F. H., & Hammer, A. L. (1994). *Applications and technical guide for the Strong Interest Inventory*. Palo Alto, CA: Consulting Psychologists Press.

Harrington, T. F. (2006). A 20-year follow-up of the Harrington-O'Shea career decision-making system. *Measurement and Evaluation in Counseling and Development*, *38*(4), 198-202.

Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, *77*(1), 81-112. doi:10.3102/003465430298487

Hawley, J. D., & Barnard, J. K. (2005). Work environment characteristics and implications for training transfer: A case study of the nuclear power industry. *Human Resource Development International*, *8*(1), 65-80. doi:10.1080/1367886042000338308

Hedges, L. V., & Nowell, A. (1995). Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science, 269*(5220)*,* 41-45. doi:10.1126/science.7604277

Hell, B., & Pässler, K. (2011). Are occupational interests hormonally influenced? The 2D:4D-interest nexus. *Personality and Individual Differences*, *51*(4), 376-380. doi:10.1016/j.paid.2010.05.033

Hell, B., Pässler, K., & Schuler, H. (2009). Was-studiere-ich.de: Konzept, Nutzen und Anwendungsmöglichkeiten [Was-studiere-ich.de: Concept, benefit, and application]. *Zeitschrift für Studium und Beratung*, *4*, 9-14.

Herold, D. M., Davis, W., Fedor, D. B., & Parsons, C. K. (2002). Dispositional influences on transfer of learning in multistage training programs. *Personnel Psychology*, *55*(4), 851-869. doi:10.1111/j.1744-6570.2002.tb00132.x

Hochholdinger, S., & Beinicke, A. (2011). Evaluation betrieblichen E-Learnings: Methoden und Befunde [Evaluation of corporate e-learning: Methods and findings]. In A. Hohenstein & K. Wilbers (Eds.), *Handbuch E-Learning*. Köln, Germany: Deutscher Wirtschaftsdienst.

Hochholdinger, S., & Beinicke, A. (2012). Potenziale und Herausforderungen netzbasierten Lernens [Potentials and challenges of net-based learning]. *Personal Quarterly*, *64*(2), 16-23.

Hochholdinger, S., & Schaper, N. (2008). Die Bedeutung von Lernkultur und Transfermotivation für betrieblichen E-Learning-Erfolg [Prediction of individual e-learning success by organizational learning culture and individual motivation to transfer]. *Zeitschrift für Personalpsychologie*, *7*(2), 81-89. doi:10.1026/1617-6391.7.2.81

Holladay, C. L., & Quiñones, M. A. (2003). Practice variability and transfer of training: The role of self-efficacy generality. *Journal of Applied Psychology*, *88*(6), 1094-1103. doi:10.1037/0021-9010.88.6.1094

Holland, J. L. (1959). A theory of vocational choice. *Journal of Counseling Psychology*, *6*(1), 35-45. doi:10.1037/h0040767

Holland, J. L. (1973). *Making vocational choices: A theory of careers*. Englewood Cliffs, NJ: Prentice-Hall.

Holland, J. L. (1985). *Making vocational choices: A theory of vocational personalities and work environments* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Holland, J. L. (1997). *Making vocational choices: A theory of vocational personalities and work environments* (3rd ed.). Odessa, FL: Psychological Assessment Resources.

Holland, J. L., Fritzsche, B., & Powell, A. (1994). *Self-directed search: Technical manual*. Odessa, FL: Psychological Assessment Resources.

Holland, J. L., & Gottfredson, G. D. (1992). Studies of the hexagonal model: An evaluation (or, the perils of stalking the perfect hexagon). *Journal of Vocational Behavior*, *40*(2), 158-170. doi:10.1016/0001-8791(92)90063-6

Holland, P. W., & Wainer, H. (Eds.). (1993). *Differential item functioning*. Hillsdale, NJ: Erlbaum.

Holton, E. F. (1996). The flawed four-level evaluation model. *Human Resource Development Quarterly*, *7*(1), 5-21. doi:10.1002/hrdq.3920070103

Holton, E. F. (2005). Holton's evaluation model: New evidence and construct elaborations. *Advances in Developing Human Resources*, *7*(1), 37-54. doi:10.1177/15234223042720 80

Holton, E. F. (2015). *Learning transfer system inventory research* [slides]. Retrieved from http://de.slideshare.net/edholton/learning-transfer-system-inventory-research

Holton, E. F., Bates, R. A., Bookter, A. I., & Yamkovenko, V. B. (2007). Convergent and divergent validity of the Learning Transfer System Inventory. *Human Resource Development Quarterly*, *18*(3), 385-419. doi:10.1002/hrdq.1210

Holton, E. F., Bates, R. A., & Ruona, W. E. A. (2000). Development of a generalized learning transfer system inventory. *Human Resource Development Quarterly, 11*(4), 333-360. doi:10.1002/1532-1096(200024)11:4<333::AID-HRDQ2>3.0.CO;2-P

Homklin, T., Takahashi, Y., & Techakanont, K. (2014, May). The influence of social and organizational support on transfer of training: Evidence from Thailand. *International Journal of Training and Development*, *18*(2), 116-131. doi:10.1111/ijtd.12031

Horn, J. L., & Noll, J. (1997). Human cognitive capabilities: Gf-Gc theory. In D. Flanagan, J. L. Genshaft, & P. Harrison (Eds.), *Contemporary intellectual assessment* (pp. 53-91). New York, NY: Guilford Press.

Hubert, L., & Arabie, P. (1987). Evaluating order hypotheses within proximity matrices. *Psychological Bulletin*, *102*(1), 172-178. doi:10.1037/0033-2909.102.1.172

Humphreys, L. G., Lubinski, D., & Yao, G. (1993). Utility of predicting group membership and the role of spatial visualization in becoming an engineer, physical scientist, or artist. *Journal of Applied Psychology, 78*(2)*,* 250-261. doi:10.1037/0021-9010.78.2.250

Hunter, J. E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. *Journal of Vocational Behavior*, *29*(3), 340-362. doi:10.1016/0001-8791(86)90013-8

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin, 96*(1)*,* 72-98. doi:10.1037/0033-2909.96.1.72

Hunter, J. E., & Schmidt, F. L. (2000). Fixed effects vs. random effects meta-analysis models: Implications for cumulative research knowledge. *International Journal of Selection and Assessment, 8*(4), 275-292. doi:10.1111/1468-2389.00156

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Thousand Oaks, CA: Sage.

Hunter, J. E., Schmidt, F. L., & Jackson, G. B. (1982). *Meta-analysis: Cumulating research findings across studies* (Vol. 4). Thousand Oaks, CA: Sage.

Hussy, W., Schreier, M., & Echterhoff, G. (2013). *Forschungsmethoden in Psychologie und Sozialwissenschaften für Bachelor* [Research methods in psychology and social sciences for the bachelor's degree]. Berlin, Germany: Springer.

Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist*, *60*(6), 581-592. doi:10.1037/0003-066x.60.6.581

Hyde, J. S., & Linn, M. C. (1988). Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin, 104*(1)*,* 53-69. doi:10.1037/0033-2909.104.1.53

IfM-Institut für Mittelstandsforschung Bonn. (2014, December 7). *KMU-Definition der Europäischen Kommission* [Definition of SME enterprises of the European Commission]. Retrieved from http://www.ifm-bonn.org/mittelstandsdefinition/definition-kmu-der-eu-kommission/

Ivanova-Stenzel, R., & Kübler, D. (2011). Gender differences in team work and team competition. *Journal of Economic Psychology*, *32*(5), 797-808. doi:10.1016/j.joep.2011.05.011

Jaccard, J., & Turrisi, R. (2003). *Interaction effects in multiple regression* (2nd ed.). Thousand Oaks, CA: Sage.

Jäger, A. O. (1984). Intelligenzstrukturforschung: Konkurrierende Modelle, neue Entwicklungen, Perspektiven [Research on intelligence structure: Competing models, new developments, perspectives]. *Psychologische Rundschau*, *35*(1), 21-35.

Jensen, A. R. (1986). Major contributions, g: Artifact or reality? *Journal of Vocational Behavior*, *29*(3), 301-331. doi:10.1016/0001-8791(86)90011-4

Johnson, W., & Bouchard, T. J. (2009). Linking abilities, interests, and sex via latent class analysis. *Journal of Career Assessment, 17*(1), 3-38. doi:10.1177/1069072708325738

Johnson, W., Carothers, A., & Deary, I. J. (2008). Sex differences in variability in general intelligence: A new look at the old question. *Perspectives on Psychological Science*, *3*(6), 518-531. doi:10.1111/j.1745-6924.2008.00096.x

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, *34*(2), 183-202. doi:10.1007/BF02289343

Jöreskog, K. G., & Sörbom, D. (1986). *LISREL VI: Analysis of linear structural relationships by the method of maximum likelihood.* Mooresville, IN: Scientific Software.

Jörin, S., Stoll, F., Bergmann, C., & Eder, F. (2003*). EXPLORIX—das Werkzeug zur Berufswahl und Laufbahnplanung. Deutschsprachige Adaptation und Weiterentwicklung des Self-Directed Search (SDS) nach John Holland (Testmanual)* [EXPLORIX—The tool for occupational choices and career planning. German adaptation and development of John Holland's Self-Directed Search (SDS)]. Bern, Switzerland: Huber.

Jörin Fux, S. (2005). *Persönlichkeit und Berufstätigkeit: Theorie und Instrumente von John Holland im deutschsprachigen Raum unter Adaptation und Weiterentwicklung von Self-directed Search (SDS) und Position Classification Inventory (PCI*) [Personality and employment: John Holland's theory and instruments in German-speaking countries and adaptation and further development of Self-directed Search (SDS) and Position Classification Inventory (PCI)] (Doctoral dissertation). Retrieved from https://cuvillier.de/uploads/preview/public_file/4784/3865377130.pdf

Judge, T. A., Jackson, C. L., Shaw, J. C., Scott, B. A., & Rich, B. L. (2007). Self-efficacy and work-related performance: The integral role of individual differences. *Journal of Applied Psychology*, *92*(1), 107-127. doi:10.1037/0021-9010.92.1.107

Kaltenbaek, J. (2003). *E-Learning und Blended-Learning in der betrieblichen Weiterbildung: Möglichkeiten und Grenzen aus Sicht von Mitarbeitern und Personalverantwortlichen in Unternehmen* [E-learning and blended-learning in corporate training: Chances and limits from the perspective of employees and T&D professionals] [data file]. Retrieved from http://www.weissensee-verlag.de/autoren/Kaltenbaek/Kaltenbaek-kurz.pdf

Kanfer, R. (1990). Motivation theory and industrial and organizational psychology. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (pp. 75-170). Palo Alto, CA: Consulting Psychologists Press.

Kanfer, R., & Ackerman, P. L. (1989). Motivation and cognitive abilities: An integrative/aptitude-treatment interaction approach to skill acquisition. *Journal of Applied Psychology*, *74*(4), 657-690. doi:10.1037/0021-9010.74.4.657

Kanfer, R., & Ackerman, P. L. (2004). Aging, adult development, and work motivation. *Academy of Management Review*, *29*(3), 440-458. doi: 10.5465/AMR.2004.13670969

*Kanfer, R., Ackerman, P. L., & Heggestad, E. D. (1996). Motivational skills and self-regulation for learning: A trait perspective. *Learning and Individual Differences, 8*(3), 185-209. doi:10.1016/S1041-6080(96)90014-X

*Kaub, K., Karbach, J., Biermann, A., Friedrich, A., Bedersdorfer, H.-W., Spinath, F. M., & Brünken, R. (2012). Berufliche Interessensorientierungen und kognitive Leistungsprofile von Lehramtsstudierenden mit unterschiedlichen Fachkombinationen [Vocational interests and cognitive ability of first-year teacher candidates as a function of selected study major]. *Zeitschrift für Pädagogische Psychologie, 26*(4), 233-249. doi:10.1024/1010-0652/a000074

Kauffeld, S. (2010). *Nachhaltige Weiterbildung. Betriebliche Seminare und Trainings entwickeln, Erfolge messen, Transfer sichern* [Sustainable training. Develop seminars and trainings, measure results, secure transfer]. doi:10.1007/978-3-540-95954-0

Kauffeld, S., Bates, R. A., Holton, E. F., & Müller, A. C. (2008). Das deutsche Lerntransfer-System-Inventar (GLTSI): Psychometrische Überprüfung der deutschsprachigen Version [The German version of the Learning Transfer Systems

Inventory (GLTSI): Psychometric validation]. *Zeitschrift für Personalpsychologie, 7*(2), 50-69. doi:10.1026/1617-6391.7.2.50

Kauffeld, S., Brennecke J., & Strack, M. (2009). Erfolge sichtbar machen: Das Maßnahmen-Erfolgs-Inventar (MEI) zur Bewertung von Trainings [Visualizing training outcomes: The MEI for training evaluations]. In S. Kauffeld, S. Grote, & E. Frieling (Eds.), *Handbuch Kompetenzentwicklung* (pp. 55-78). Stuttgart, Germany: Schäffer-Poeschel.

Kaufman, A. S., Johnson, C. K., & Xin Liu (2008). A CHC theory-based analysis of age differences on cognitive abilities and academic skills at ages 22 to 90 years. *Journal of Psychoeducational Assessment, 26*(4)*,* 350-381. doi:10.1177/0734282908314108

Keith, N., & Frese, M. (2008). Effectiveness of error management training: A meta-analysis. *Journal of Applied Psychology*, *93*(1), 59-69. doi:10.1037/0021-9010.93.1.59

*Kelso, G. I., Holland, J. L., & Gottfredson, G. D. (1977). The relation of self-reported competencies to aptitude test scores. *Journal of Vocational Behavior, 10*(1), 99-103. doi:10.1016/0001-8791(77)90046-X

Kenny, D. A., Kaniskan, B., & McCoach, D. B. (2014). The performance of RMSEA in models with small degrees of freedom. *Sociological Methods & Research*. Advance online publication. doi:10.1177/0049124114543236

Kepes, S., McDaniel, M. A., Brannick, M. T., & Banks, G. C. (2013). Meta-analytic reviews in the organizational sciences: Two meta-analytic schools on the way to MARS (the Meta-analytic Reporting Standards). *Journal of Business and Psychology, 28*(2), 123-143. doi:10.1007/s10869-013-9300-2

Kerres, M., & De Witt, C. (2003). A didactical framework for the design of blended learning arrangement. *Journal of Educational Media*, *28*(2-3), 101-113. doi:10.1080/1358165032000165653

Khasawneh, S., Bates, R. A., & Holton, E. F. (2006). Construct validation of an Arabic version of the Learning Transfer System Inventory for use in Jordan. *International Journal of Training and Development*, *10*(3), 180-194. doi:10.1111/j.1468-2419.2006.00253.x

*Kirchler, E. (1990). *Reintegration von Arbeitslosen* [Reintegration of unemployed people] [Research report]. University of Linz, Austria: Department of Educational Science and Psychology.

Kirkpatrick Partners (2014, December 7). *The Kirkpatrick Model*. Retrieved from http://www.kirkpatrickpartners.com/OurPhilosophy/TheKirkpatrickModel/tabid/302/Default.aspx

Kirkpatrick, D. L. (1959). Techniques for evaluating training programs. *Journal of American Society of Training Directors*, *13*(3), 21-26.

Kirkpatrick, D. L. (1967). Evaluation of training. In R. Craig & L. Bittel (Eds.), *Training and development handbook* (pp. 87-112). New York, NY: McGraw-Hill.

Kirkpatrick, D. L. (1975). Techniques for evaluating training programs. In D. L. Kirkpatrick (Ed.), *Evaluating training programs*. Alexandria, VA: ASTD.

Kirkpatrick, D. L. (1994). *Evaluating training programs: The four levels*. San Francisco, CA: Berrett-Koehler.

Klimoski, R. J. (1991). Theory presentation in human resource management. *Human Resource Management Review*, *1*(4), 253-271. doi:10.1016/1053-4822(91)90002-T

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, *119*(2), 254-284. doi:10.1037/0033-2909.119.2.254

Kontoghiorghes, C. (2001). Factors affecting training effectiveness in the context of the introduction of new technology: A US case study. *International Journal of Training and Development*, *5*(4), 248-260. doi:10.1111/1468-2419.00137

Kontoghiorghes, C. (2004). Reconceptualizing the learning transfer conceptual framework: Empirical validation of a new systemic model. *International Journal of Training and Development*, *8*(3), 210-221. doi:10.1111/j.1360-3736.2004.00209.x

Kraiger, K. (2003). Perspectives on training and development. In W. C. Borman, D. R. Ilgen, & R. J. Klimoski (Eds.), *Handbook of psychology: Industrial and organizational psychology* (Vol. 12, pp. 171-192). doi:10.1002/0471264385.wei1208

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*(2), 311-328. doi:10.1037/0021-9010.78.2.311

*Krapic, N., Kardum, I., & Kristofic, B. (2008). Relationship between personality traits, intelligence and vocational interests. *Psihologijske Teme, 17*(1), 75-91.

Kumkale, G. T., & Albarracín, D. (2004). The sleeper effect in persuasion: A meta-analytic review. *Psychological Bulletin*, *130*(1), 143-172. doi:10.1037/0033-2909.130.1.143

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2004). Academic performance, career potential, creativity, and job performance: Can one construct predict them all? *Journal of Personality and Social Psychology, 86*(1)*,* 148-161. doi:10.1037/0022-3514.86.1.148

Kurtz-Costes, B. E., & Schneider, W. (1994). Self-concept, attributional beliefs, and school achievement: A longitudinal analysis. *Contemporary Educational Psychology*, *19*(2), 199-216. doi:10.1006/ceps.1994.1017

Landis, R. S. (2013). Successfully combining meta-analysis and structural equation modeling: Recommendations and strategies. *Journal of Business Psychology*, *28*(3), 251-261. doi:10.1007/s10869-013-9285-x

Larson, L. M., Rottinghaus, P. J., & Borgen, F. H. (2002). Meta-analyses of big six interests and big five personality factors. *Journal of Vocational Behavior, 61*(2)*,* 217-239. doi:10.1006/jvbe.2001.1854

Latham, G. P. (2007). *Work motivation. History, theory, research, and practice*. Thousand Oaks, CA: Sage.

Lee, Y., Driscoll, M. P., & Nelson, D. W. (2004). The past, present, and future of research in distance education: Results of a content analysis. *American Journal of Distance Education*, *18*(4), 225-241. doi:10.1207/s15389286ajde1804_4

Leiner, D. J. (2013). *SoSci Survey (oFb–der onlineFragebogen)* (Version 2.3.05-i) [computer software]. Available from https://www.soscisurvey.de

Lemos, G. C., Abad, F. J., Almeida, L. S., & Colom, R. (2013). Sex differences on g and non-g intellectual performance reveal potential sources of STEM discrepancies. *Intelligence, 41*(1)*,* 11-18. doi:10.1016/j.intell.2012.10.009

Lenhard, W., & Lenhard, A. (2014). *Calculation of effect sizes* [computer software]. Available from http://www.psychometrica.de/effektstaerke.html

Lenske, W., & Werner, D. (2009). *Umfang, Kosten und Trends der betrieblichen Weiterbildung–Ergebnisse der IW-Weiterbildungserhebung 2008* [Trends and costs of further training in Germany–Results of the IW survey 2008]. doi:10.2373/1864-810X.09-01-04

Lent, R. W., Brown, S. D., & Hackett, G. (1994). Toward a unifying social cognitive theory of career and academic interest, choice, and performance. *Journal of Vocational Behavior*, *45*(1), 79-122. doi:10.1006/jvbe.1994.1027

Leung, S. A. (2008). The big five career theories. In J. A. Athanasou & R. Van Esbroeck (Eds.), *International handbook of career guidance* (pp. 115-132). doi:10.1007/978-1-4020-6230-8_6

Lim, D. H., & Johnson, S. D. (2002). Trainee perceptions of factors that influence learning transfer. *International Journal of Training and Development*, *6*(1), 36-48. doi:10.1111/1468-2419.00148

Lindberg, S. M., Hyde, J. S., Petersen, J. L., & Linn, M. C. (2010). New trends in gender and mathematics performance: A meta-analysis. *Psychological Bulletin, 136*(6)*,* 1123-1135. doi:10.1037/a0021276

Linn, R. L. (1978). Single-group validity, differential validity, and differential prediction. *Journal of Applied Psychology*, *63*(4), 507-512. doi:10.1037/0021-9010.63.4.507

Lippa, R. A. (1998). Gender-related individual differences and the structure of vocational interests: The importance of the people-things dimension. *Journal of Personality and Social Psychology, 74*(4)*,* 996-1009. doi:10.1037/0022-3514.74.4.996

Lippa, R. A. (2010). Gender differences in personality and interests: When, where, and why? *Social and Personality Psychology Compass*, *4*(11), 1098-1110. doi:10.1111/j.1751-9004.2010.00320.x

Liu, I. M., & Agresti, A. (1996). Mantel-Haenszel-type inference for cumulative odds ratios with a stratified ordinal response. *Biometrics*, *52(*4), 1223-1234. doi:10.2307/2532838

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist*, *57*(9), 705-717. doi:10.1037/0003-066X.57.9.705

Long, L., & Tracey, T. J. (2006). Structure of RIASEC scores in China: A structural meta-analysis. *Journal of Vocational Behavior*, *68*(1), 39-51. doi:10.1016/j.jvb.2005.01.002

Low, K. S., & Rounds, J. (2007). Interest change and continuity from early adolescence to middle adulthood. *International Journal for Educational and Vocational Guidance, 7*(1)*,* 23-36. doi:10.1007/s10775-006-9110-4

Low, K. S., Yoon, M., Roberts, B. W., & Rounds, J. (2005). The stability of vocational interests from early adolescence to middle adulthood: A quantitative review of longitudinal studies. *Psychological Bulletin, 131*(5)*,* 713-737. doi:10.1037/0033-2909.131.5.713

Lowman, R. L. (1991). *The clinical practice of career assessment: Interests, abilities, and personality*. Washington, DC: American Psychological Association.

Lowman, R. L. (2010). Interests. In I. B. Weiner & W. E. Craighead (Eds.), *The Corsini Encyclopedia of Psychology* (4th ed., Vol. 2, pp. 844-846). Hoboken, NJ: Wiley.

Lowman, R. L., & Carson, A. D. (2013). Conceptualization and assessment of interests. In J. Graham & J. Naglieri (Eds.), *Handbook of psychology: Assessment psychology* (2nd ed., pp. 534-557). New York, NY: Wiley.

Lowman, R. L., & Leeman, G. E. (1988). The dimensionality of social intelligence: Social abilities, interests, and needs. *Journal of Psychology, 122*(3), 279-290. doi:10.1080/00223980.1988.9915516

*Lowman, R. L., Williams, R. E., & Leeman, G. E. (1985). The structure and relationship of college women's primary abilities and vocational interests. *Journal of Vocational Behavior, 27*(3), 298-315. doi:10.1016/0001-8791(85)90038-7

Lubinski, D. (2000). Scientific and social significance of assessing individual differences: Sinking shafts at a few critical points. *Annual Review of Psychology, 51,* 405-444. doi:10.1146/annurev.psych.51.1.405

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, *1*(2), 130-149. doi:10.1037/1082-989X.1.2.130

Machin, M. A., & Fogarty, G. J. (1997). The effects of self-efficacy, motivation to transfer, and situational constraints on transfer intentions and transfer of training. *Performance Improvement Quarterly*, *10*(2), 98-115. doi:10.1111/j.1937-8327.1997.tb00051.x

Machtmes, K., & Asher, J. W. (2000). A meta-analysis of the effectiveness of telecourses in distance education. *American Journal of Distance Education*, *14*(1), 27-46. doi:10.1080/08923640009527043

Mackintosh, N. J. (2011). *IQ and human intelligence* (2nd ed.). Oxford, UK: Oxford University Press.

Macpherson, A., Elliot, M., Harris, I., & Homan, G. (2004). E-learning: Reflections and evaluation of corporate programmes. *Human Resource Development International*, *7*(3), 295-313. doi:10.1080/13678860310001630638

Mandl, H., Prenzel, M., & Gräsel, C. (1992). Das Problem des Lerntransfers in der betrieblichen Weiterbildung [The problem of learning transfer in personnel training]. *Unterrichtswissenschaft,* 20, 126-143.

*Marcus, B., Wagner, U., Poole, A., Powell, D. M., & Carswell, J. (2009). The relationship of GMA to counterproductive work behavior revisited. *European Journal of Personality, 23*(6), 489-507. doi:10.1002/per.728

Markóczy, L. (1997). Measuring beliefs: Accept no substitutes. *Academy of Management Journal*, *40*(5), 1228-1242. doi:10.2307/256934

Marler, J. H., Liang, X., & Dulebohn, J. H. (2006). Training and effective employee information technology use. *Journal of Management*, *32*(5), 721-743. doi:10.1177/0149206306292388

Martocchio, J. J. (1994). Effects of conceptions of ability on anxiety, self-efficacy, and learning in training. *Journal of Applied Psychology*, *79*(6), 819-825. doi:10.1037/0021-9010.79.6.819

Martocchio, J. J., & Judge, T. A. (1997). Relationship between conscientiousness and learning in employee training: Mediating influences of self-deception and self-efficacy. *Journal of Applied Psychology*, *82*(5), 764-773. doi:10.1037/0021-9010.82.5.764

Martocchio, J. J., & Webster, J. (1992). Effects of feedback and cognitive playfulness on performance in microcomputer software training. *Personnel Psychology*, *45*(3), 553-578. doi:10.1111/j.1744-6570.1992.tb00860.x

McCrae, R. R., & Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology*, *52*(1), 81-90. doi:10.1037/0022-3514.52.1.81

Meade, A. W., & Tonidandel, S. (2010). Not seeing clearly with Cleary: What test bias analyses do and do not tell us. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, *3*(2), 192-205. doi:10.1111/j.1754-9434.2010.01223.x

Means, B., Toyama, Y., Murphy, R. F., & Baki, M. (2013). The effectiveness of online and blended learning: A meta-analysis of the empirical literature. *Teachers College Record*, *115*(3), 1-47.

Means, B., Toyama, Y., Murphy, R. F., Baki, M., & Jones, K. (2010). *Evaluation of evidence-based practices in online learning: A meta-analysis and review of online learning studies.* (Research Report No. ED-04-CO-0040 Task 0006) [data file]. Retrieved from U.S. Department of Education, Office of Planning, Evaluation, and Policy Development website: http://www2.ed.gov/rschstat/eval/tech/evidence-based-practices/

finalreport.pdf

Milliken, F. J., & Martins, L. L. (1996). Searching for common threads: Understanding the multiple effects of diversity in organizational groups. *Academy of Management Review*, *21*(2), 402-433. doi:10.5465/AMR.1996.9605060217

Mitchell, L. K., & Krumboltz J. D. (1996). Krumboltz's learning theory of career choice and counseling. In D. Brown & L. Brooks (Eds.), *Career choice and development* (3rd ed., pp. 233-280). San Francisco, CA: Jossey-Bass.

MMB-Institut für Medien- und Kompetenzforschung. (2010). *Telefonische Befragung zum Einsatz von eLearning in deutschen Großunternehmen* [Telephone survey on the implementation of e-learning in large German companies] [data file]. Retrieved from http://www.mmb-institut.de/projekte/digitales-lernen/Einsatz-von-E-Learning-in-deutschen-Grossunternehmen.pdf

MMB-Institut für Medien und Kompetenzforschung. (2012). *Mobile Learning–kurzer Hype oder stabiler Megatrend?* [Mobile Learning–short hype or stable megatrend?] [data file]. Retrieved from http://www.mmb-institut.de/monitore/trendmonitor/MMB-Trendmonitor_2012_I.pdf

Moosbrugger, H., & Kelava, A. (2012). *Testtheorie und Fragebogenkonstruktion* [Test theory and questionnaire design]. Berlin, Germany: Springer.

*Mussel, P. (2013). Intellect: A theoretical framework for personality traits related to intellectual achievements. *Journal of Personality and Social Psychology, 104*(5), 885-906. doi:10.1037/a0031918

Nagy, G. (2006). *Berufliche Interessen, kognitive und fachgebundene Kompetenzen: Ihre Bedeutung für die Studienfachwahl und die Bewährung im Studium* [Vocational interests, cognitive and scholastic abilities: Their role in choice of major and success at university] (Doctoral dissertation). Retrieved from http://www.diss.fu-berlin.de/2007/109/

Nagy, G., Trautwein, U., & Lüdtke, O. (2010). The structure of vocational interests in Germany: Different methodologies, different conclusions. *Journal of Vocational Behavior*, *76*(2), 153-169. doi:10.1016/j.jvb.2007.07.002

National Institute of Education (NIE). (1975). Guidelines for the assessment of sex bias and sex fairness in career interest inventories. *Measurement and Evaluation in Guidance*, *8*(1), 7-11.

Newton, R., & Doonga, N. (2007). Corporate e-learning: Justification for implementation and evaluation of benefits. A study examining the views of training managers and training providers. *Education for Information*, *25*(2), 111-130.

Noe, R. A. (1986). Trainee attributes and attitudes: Neglected influences on training effectiveness. *Academy of Management Review*, *11*(4), 736-749. doi:10.2307/258393

Noe, R. A., & Schmitt, N. (1986). The influence of trainee attitudes on training effectiveness: Test of a model. *Personnel Psychology*, *39*(3), 497-523. doi:10.1111/j.1744-6570.1986.tb00950.x

Nye, C. D., Su, R., Rounds, J., & Drasgow, F. (2012). Vocational interests and performance: A quantitative summary of over 60 years of research. *Perspectives on Psychological Science*, *7*(4), 384-403. doi:10.1177/1745691612449021

Orvis, K. A., Brusso, R. C., Wasserman, M. E., & Fisher, S. L. (2010). E-nabled for e-learning? The moderating role of personality in determining the optimal degree of learner control in an e-learning environment. *Human Performance*, *24*(1), 60-78. doi:10.1080/08959285.2010.530633

Orvis, K. A., Fisher, S. L., & Wasserman, M. E. (2009). Power to the people: Using learner control to improve trainee reactions and learning in web-based instructional environments. *Journal of Applied Psychology*, *94*(4), 960-971. doi:10.1037/a0014977

Osipow, S. H. (1999). Assessing career indecision. *Journal of Vocational Behavior*, *55*(1), 147-154. doi:10.1006/jvbe.1999.1704

Osterlind, S. J., & Everson, H. T. (2009). *Differential item functioning* (2nd ed.). Thousand Oaks, CA: Sage.

Park, J.-H., & Wentling, T. (2007). Factors associated with transfer of training in workplace e-learning. *Journal of Workplace Learning*, *19*(5), 311-329. doi:10.1108/13665620710757860

Pässler, K. (2011). *Die Bedeutung beruflicher Interessen und kognitiver Fähigkeiten für die Studien- und Berufswahl* [The importance of vocational interests and cognitive abilities for educational and occupational choice] (Doctoral dissertation). Retrieved from http://opus.uni-hohenheim.de/volltexte/2012/699/

Pässler, K., Beinicke, A., & Hell, B. (2014). Gender-related differential validity and differential prediction in interest inventories. *Journal of Career Assessment*, *22*(1), 138-152. doi:10.1177/1069072713492934

Pässler, K., Beinicke, A., & Hell, B. (2015). Interests and intelligence: A meta-analysis. *Intelligence*, *50*, 30-51. doi:10.1016/j.intell.2015.02.001

*Pässler, K., & Hell, B. (2012). Do interests and cognitive abilities help explain college major choice equally well for women and men? *Journal of Career Assessment, 20*(4), 479-496. doi:10.1177/1069072712450009

Patel, L. (2010). *ASTD State of the industry report 2010* (Report No. 191001). Alexandria, VA: American Society for Training and Development Press.

Patton, W., & McMahon, M. (2006). *Career development and systems theory: Connecting theory and practice* (2nd ed.). Rotterdam, Netherlands: Sense.

Paustian-Underdahl, S. C. (2012). *A meta-analysis of gender differences in leadership effectiveness* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3539035)

Penfield, R. D. (2009). DIFAS: Differential item functioning analysis system (version 5.0) [computer software]. Available from http://www.education.miami.edu/facultysites/penfield/

Penfield, R. D., & Algina, J. (2006). A generalized DIF effect variance estimator for measuring unsigned differential test functioning in mixed format tests. *Journal of Educational Measurement*, *43*(4), 295-312. doi:10.1111/j.1745-3984.2006.00018.x

Ployhart, R. E., & Turner, S. F. (2014). Organizational adaptability. In D. Chan (Ed.), *Individual adaptability to changes at work: New directions in research* (pp. 73-91). New York, NY: Routledge.

Podsakoff, P. M., MacKenzie, S. B., Lee, J., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*(5), 879-903. doi:10.1037/0021-9010.88.5.879

Pollard, E., & Hillage J. (2001). *Exploring e-learning* (Report No. 376). Retrieved from University of Sussex, Institute for Employment Studies website: http://www.employment-studies.co.uk/summary/summary.php?id=376

Prediger, D. J., & Cole, N. S. (1975). Sex-role socialization and employment realities: Implications for vocational-interest measures. *Journal of Vocational Behavior*, *7*(2), 239-251. doi:10.1016/0001-8791(75)90064-0

Prediger, D. J., & Lamb, R. R. (1979). The validity of sex-balanced and sex-restrictive vocational interest reports: A comparison. *Vocational Guidance Quarterly*, *28*(1), 16-24. doi:10.1002/j.2164-585x.1979.tb00079.x

*Proyer, R. T. (2006). The relationship between vocational interests and intelligence: Do findings generalize across different assessment methods? *Psychology Science, 48*(4), 463-476.

Pugh, K. J., & Bergin, D. A. (2006). Motivational influences on transfer. *Educational Psychologist*, *41*(3), 147-160. doi:10.1207/s15326985ep4103_2

Raju, N. S., Burke, M. J., Normand, J., & Langlois, G. M. (1991). A new meta-analytic approach. *Journal of Applied Psychology, 76*(3), 432-446. doi:10.1037/0021-9010.76.3.432

Raju, N. S., & Fleer, P. F. (2003). *A computer program for conducting validity generalization analysis.* Chicago, IL: Institute of Technology.

Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality*, *41*(1), 203-212. doi:10.1016/j.jrp.2006.02.001

*Randahl, G. J. (1991). A typological analysis of the relations between measured vocational interests and abilities. *Journal of Vocational Behavior, 38*(3), 333-350. doi:10.1016/0001-8791(91)90034-J

Reardon, R. C., Bullock, E. E., & Meyer, K. E. (2007). A Holland perspective on the U.S. workforce from 1960 to 2000. *The Career Development Quarterly, 55*(3), 262-274. doi:10.1002/j.2161-0045.2007.tb00082.x

Ree, M. J., & Earles, J. A. (1991). Predicting training success: Not much more than g. *Personnel Psychology*, *44*(2), 321-332. doi:10.1111/j.1744-6570.1991.tb00961.x

*Reeve, C. L., & Heggestad, E. D. (2004). Differential relations between general cognitive ability and interest-vocation fit. *Journal of Occupational and Organizational Psychology, 77*(3), 385-402. doi:10.1348/0963179041752673

Rey, G. D. (2009). *E-Learning. Theorien, Gestaltungsempfehlungen und Forschung* [E-learning. Theories, recommendations for design, and research]. Bern, Schweiz: Huber.

Reynolds, C. R., Vannest, K. J., & Fletcher-Janzen, E. (Eds.). (2013). *Encyclopedia of Special Education*. Hoboken, NJ: Wiley.

Robbins, S. P., & Judge, T. A. (2009). *Organizational behavior* (13th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

*Rolfhus, E. L., & Ackerman, P. L. (1996). Self-report knowledge: At the crossroads of ability, interest, and personality. *Journal of Educational Psychology, 88*(1), 174-188. doi:10.1037/0022-0663.88.1.174

Rosenberg, M. J. (1995). Performance technology, performance support, and the future of training: A commentary. *Performance Improvement Quarterly*, *8*(1), 94-99. doi:10.1111/j.1937-8327.1995.tb00662.x

Rossett, A., & Marshall, J. (2010). What corporate training professionals think about e-learning: Practitioners' views on the potential of e-learning in the workplace. *Journal of Asynchronous Learning Networks*, *14*(2), 19-27. Retrieved from http://files.eric.ed.gov/fulltext/EJ909870.pdf

Rottinghaus, P. J., Coon, K. L., Gaffey, A. R., & Zytowski, D. G. (2007). Thirty-year stability and predictive validity of vocational interests. *Journal of Career Assessment*, *15*, 5-22. doi:10.1177/1069072706294517

Rouiller, J. Z., & Goldstein, I. L. (1993). The relationship between organizational transfer climate and positive transfer of training. *Human Resource Development Quarterly*, *4*(4), 377-390. doi:10.1002/hrdq.3920040408

Rounds, J., & Day, S. X. (1999). Describing, evaluating, and creating vocational interest structures. In M. L. Savickas & A. R. Spokane (Eds.), *Vocational interests: Meaning measurement, and counseling use* (pp. 103-133). Palo Alto, CA: Davies-Black.

Rounds, J., & Tracey, T. J. (1996). Cross-cultural structural equivalence of RIASEC models and measures. *Journal of Counseling Psychology*, *43*(3), 310-329. doi:10.1037/0022-0167.43.3.310

Rounds, J., Tracey, T. J., & Hubert, L. (1992). Methods for evaluating vocational interest structural hypotheses. *Journal of Vocational Behavior*, *40*(2), 239-259. doi:10.1016/0001-8791(92)90073-9

Ruble, D. N., Martin, C. L., & Berenbaum, S. A. (2006). Gender development. In N. Eisenberg, W. Damon, R. M. Lerner, N. Eisenberg, W. Damon, & R. M. Lerner (Eds.), *Handbook of child psychology* (6th ed., Vol. 3, pp. 858-932). Hoboken, NJ: Wiley.

Ruona, W. E. A., Leimbach, M., Holton, E. F., & Bates, R. A. (2002). The relationship between learner utility reactions and predicted learning transfer among trainees. *International Journal of Training and Development*, *6*(4), 218-228. doi:10.1111/1468-2419.00160

Russell, M. (2007). *Assessing vocational interests: Convergence and divergence of inventories and informants* (Unpublished doctoral dissertation). University of Illinois at Urbana-Champaign, Urbana, IL.

Saks, A. M. (1997). Transfer of training and self-efficacy: What is the dilemma? *Applied Psychology*, *46*(4), 365-370. doi:10.1111/j.1464-0597.1997.tb01240.x

Saks, A. M. (2002). So what is a good transfer of training estimate? A reply to Fitzpatrick. *The Industrial-Organizational Psychologist*, *39*(3), 29-30. doi:10.1037/e576922011-004

Saks, A. M., & Burke, L. A. (2012). An investigation into the relationship between training evaluation and the transfer of training. *International Journal of Training and Development, 16*(2), 118-127. doi:10.1111/j.1468-2419.2011.00397.x

Saks, A. M., & Burke-Smalley, L. A. (2014). Is transfer of training related to firm performance? *International Journal of Training and Development*, *18*(2), 104-115. doi:10.1111/ijtd.12029

Saks, A. M., Salas, E., & Lewis, P. (2014). The transfer of training. *International Journal of Training and Development, 18*(2), 81-83. doi:10.1111/ijtd.12032

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*(2), 74-101. doi:10.1177/1529100612436661

Salas, E., Wilson, K. A., Priest, H. A., & Guthrie, J. W. (2006). Design, delivery, and evaluation of training systems. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (3rd ed., pp. 472-512). doi:10.1002/0470048204.ch18

Sandelands, E., & Wills, M. (1996). Creating virtual support for lifelong learning. *The Learning Organization, 3*(5), 26-31. doi:10.1108/09696479610131215

Savickas, M. L. (2004). The theory and practice of career construction. In S. D. Brown & R. W. Lent (Eds.), *Career development and counseling: Putting theory and research to work* (1st ed., pp. 42-70). Hoboken, NJ: Wiley.

Savickas, M. L. (2006). Career construction theory. In J. H. Greenhaus & G. A. Callanan (Eds.), *Encyclopedia of career development* (pp. 85-89). Thousand Oaks, CA: Sage. doi:10.4135/9781412952675.n34

Savickas, M. L. (2011). *Career counseling*. Washington, DC: American Psychological Association.

Savickas, M. L. (2013). Career construction theory and practice. In S. D. Brown & R. W. Lent (Eds.), *Career development and counseling: Putting theory and research to work* (2nd ed., pp. 147-187). Hoboken, NJ: Wiley.

Savickas, M. L., Nota, L., Rossier, J., Dauwalder, J.-P., Duarte, M. E., Guichard, J., … Van Vianen, A. E. M. (2009). Life designing: A paradigm for career construction in the 21st century. *Journal of Vocational Behavior*, *75*(3), 239-250. doi:10.1016/j.jvb.2009.04.004

Schaie, K. W. (2013). *Developmental influences on adult intelligence: The Seattle longitudinal study* (2nd ed.). Oxford, UK: Oxford University Press.

Schlenker, B. R. (1980). *Impression management: The self-concept, social identity, and interpersonal relations* (pp. 21-43). Monterey, CA: Brooks/Cole Publishing Company.

Schlenker, B. R. (2003). Self-presentation. In M. R. Leary & J. P. Tangney (Eds.), *Handbook of self and identity* (pp. 492-518). New York, NY: Guilford Press.

*Schmidt, D. B., Lubinski, D., & Benbow, C. P. (1998). Validity of assessing educational-vocational preference dimensions among intellectually talented 13-year-olds. *Journal of Counseling Psychology, 45*(4), 436-453. doi:10.1037/0022-0167.45.4.436

Schmidt, F. L. (2011). A theory of sex differences in technical aptitude and some supporting evidence. *Perspectives on Psychological Science, 6*(6)*,* 560-573. doi:10.1177/1745691611419670

Schmidt, F. L. (2014). A general theoretical integrative model of individual differences in interests, abilities, personality traits, and academic and occupational achievement: A commentary on four recent articles. *Perspectives on Psychological Science, 9*(2)*,* 211-218. doi:10.1177/1745691613518074

Schmidt, F. L., & Hunter, J. (2004). General mental ability in the world of work: Occupational attainment and job performance. *Journal of Personality and Social Psychology, 86*(1)*,* 162-173. doi:10.1037/0022-3514.86.1.162

Schneider, W. (2000). Giftedness, expertise, and (exceptional) performance: A developmental perspective. In K. A. Heller, F. J. Mönks, R. J. Sternberg, & R. F. Subotnik (Eds.), *International handbook of giftedness and talent* (pp. 165-177). Oxford, UK: Pergamon.

Schneider, W. (2008). Expertiseerwerb [Acquisition of expertise]. In W. Schneider & M. Hasselhorn (Eds.), *Handbuch der pädagogischen psychologie* (pp. 136-144). Göttingen, Germany: Hogrefe.

Schneider, W., & McGrew, K. S. (2012). The Cattell-Horn-Carroll model of intelligence. In D. Flanagan & P. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 99-144). New York, NY: Guilford Press.

Schriver, R., & Giles, S. (1999). Real ROI numbers. *Training and Development*, *53*(8), 51-52.

Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagne, & M. Scriven (Eds.), *Perspectives of curriculum evaluation* (pp. 39-83). Chicago, IL: Rand McNally.

Seyda, S., & Werner, D. (2012). *IW-Weiterbildungserhebung 2011–Gestiegenes Weiterbildungsvolumen bei konstanten Kosten* [IW-continuous training survey 2011–Increasing involvement at constant costs]. doi:10.2373/1864-810X.12-01-03

Seyler, D. L., Holton, E. F., Bates, R. A., Burnett, M. F., & Carvalho, M. B. (1998). Factors affecting motivation to transfer training. *International Journal of Training and Development*, *2*(1), 2-16. doi:10.1111/1468-2419.00031

Shackel, B. (2009). Usability–context, framework, definition, design and evaluation. *Interacting with Computers*, *21*(5-6), 339-346. doi:10.1016/j.intcom.2009.04.007

Shmikler, S. (1996). *Promoting training transfer with technology based learning.* Paper presented at the Computer Support and Training Conference, Nashville.

Sitzmann, T., Brown, K. G., Ely, K., Kraiger, K., & Wisher, R. A. (2009). A cyclical model of motivational constructs in web-based courses. *Military Psychology*, *21*(4), 534-551. doi:10.1080/08995600903206479

Sitzmann, T., Kraiger, K., Stewart, D., & Wisher, R. A. (2006). The comparative effectiveness of web-based and classroom instruction: A meta-analysis. *Personnel Psychology*, *59*(3), 623-664. doi:10.1111/j.1744-6570.2006.00049.x

Smith, S. M. (1994). Theoretical principles of context-dependent memory. In M. Gruneberg & P. E. Morris (Eds.), *Theoretical aspects of memory* (2nd ed., pp. 168-195). doi:10.4324/9780203978108

Smith, S. M., & Vela, E. (2001). Environmental context-dependent memory: A review and meta-analysis. *Psychonomic Bulletin & Review*, *8*(2), 203-220. doi:10.3758/BF03196157

Smith-Jentsch, K. A., Cannon-Bowers, J. A., Tannenbaum, S. I., & Salas, E. (2008). Guided team self-correction: Impacts on team mental models, processes, and effectiveness. *Journal of Small Group Research*, *39*(3), 303-327. doi:10.1177/1046496408317794

Solga, M. (2011a). Förderung von Lerntransfer [Promoting learning transfer]. In J. Ryschka, M. Solga, & M. Mattenklott (Eds.), *Praxishandbuch Personalentwicklung. Instrumente, Konzepte, Beispiele* (pp. 339-368). doi:10.1007/978-3-8349-6384-0_6

Solga, M. (2011b). Evaluation der Personalentwicklung [Evaluation of human resource development]. In J. Ryschka, M. Solga, & M. Mattenklott (Eds.), *Praxishandbuch Personalentwicklung Instrumente, Konzepte, Beispiele* (pp. 369-399). doi:10.1007/978-3-8349-6384-0_7

Spector, P. E. (2006). Method variance in organizational research: Truth or urban legend? *Organizational Research Methods*, *9*(2), 221-232. doi:10.1177/1094428105284955

Spitz-Oener, A. (2006). Technical change, job tasks, and rising educational demands: Looking outside the wage structure. *Journal of Labor Economics, 24*(2), 235-270. doi:10.1086/499972

Spokane, A. R., Meir, E. I., & Catalano, M. (2000). Person-environment congruence and Holland's theory. *Journal of Vocational Behavior*, *57*(2), 137-187. doi:10.1006/jvbe.2000.1771

Staggs, G. D., Larson, L. M., & Borgen, F. H. (2007). Convergence of personality and interests: Meta-analysis of the multidimensional personality questionnaire and the strong interest inventory. *Journal of Career Assessment, 15*(4), 423-445. doi:10.1177/1069072707305760

Stankov, L. (2000). The theory of fluid and crystallized intelligence. *Learning and Individual Differences, 12*(1), 1-3. doi:10.1016/S1041-6080(00)00039-X

*Stanley, J., Stumpf, H., & Cohn, S. (1995). Ipsative evaluative attitudes versus vocational interests and cognitive abilities in bright male and female seventh-graders. In N. Colangelo, S. Assouline, & D. L. Ambroson (Eds.), *Talent Development* (Vol. 3, pp. 41-58). Dayton, OH: Ohio Psychology Press.

Steiner, D. D., Dobbins, G. H., & Trahan, W. A. (1991). The trainer-trainee interaction: An attributional model of training. *Journal of Organizational Behavior*, *12*(4), 271-286. doi:10.1002/job.4030120403

Strand, S., Deary, I. J., & Smith, P. (2006). Sex differences in cognitive abilities test scores: A UK national picture. *The British Journal of Educational Psychology, 76*(3), 463-480. doi:10.1348/000709905X50906

Strong, E. K. (1943). *Vocational interests of men and women*. Palo Alto, CA: Stanford University Press.

Su, R., Rounds, J., & Armstrong, P. I. (2009). Men and things, women and people: A meta-analysis of sex differences in interests. *Psychological Bulletin*, *135*(6), 859-884. doi:10.1037/a0017364

Super, D. E. (1969). Vocational development theory: Persons, positions, and processes. *The Counseling Psychologist, 1*(1), 2-9. doi:10.1177/001100006900100101

Super, D. E. (1980). A life-span, life-space approach to career development. *Journal of Vocational Behavior, 16*(3), 282-298. doi:10.1016/0001-8791(80)90056-1

Super, D. E. (1990). A life-span, life-space approach to career development. In D. Brown & L. Brooks (Eds.), *Career choice and development: Applying contemporary theories to practice* (2nd ed., pp. 197-261). San Francisco, CA: Jossey-Bass.

Swanson, J. L. (1992). The structure of vocational interests for African-American college students. *Journal of Vocational Behavior*, *40*(2), 129-143. doi:10.1016/0001-8791(92)90062-5

Swanson, J. L. (1999). Stability and change in vocational interests. In M. L. Savickas & A. R. Spokane (Eds.), *Vocational interests: Meaning, measurement, and counseling use* (pp. 135-158). Palo Alto, CA: Davies-Black.

Swanson, J. L. (2013). Traditional and emerging career development theory and the psychology of working. In D. L. Blustein (Ed.), *The Oxford handbook of the psychology of working* (pp. 49-67). doi:10.1093/oxfordhb/9780199758791.013.0004

Swanson, J. L., & Gore, P. A. (2000). Advances in vocational psychology theory and research. In S. D. Brown & R. W. Lent (Eds.), *Handbook of counseling psychology* (3rd ed., pp. 233-269). Hoboken, NJ: Wiley.

Swanson, S. (2001). E-learning branches out. *InformationWeek*, 42-60. Retrieved from http://www.informationweek.com/826/elearning.htm

Tai, L. (2005). *Corporate e-learning: How e-learning is created in three large corporations* (Doctoral Dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3168048)

Tan, J. A., Hall, R. J., & Boyce, C. (2003). The role of employee reactions in predicting training effectiveness. *Human Resource Development Quarterly*, *14*(4), 397-411. doi:10.1002/hrdq.1076

Tannenbaum, S. I., Smith-Jentsch, K. A., & Behson, S. J. (1998). Training team leaders to facilitate team learning and performance. In J. A. Cannon-Bowers & E. Salas (Eds.), *Decision making under stress: Implications for training and simulation* (pp. 247-270). Washington, DC: APA Press.

Tannenbaum, S. I., & Yukl, G. (1992). Training and development in work organizations. *Annual Review of Psychology*, *43*(1), 399-441. doi:10.1146/annurev.ps.43.020192.002151

Taylor, P. J., Russ-Eft, D. F., & Chan, D. W. L. (2005). A meta-analytic review of behavior modeling training. *Journal of Applied Psychology*, *90*(4), 692-709. doi:10.1037/0021-9010.90.4.692

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*(3), 251-273. doi:10.1016/j.hrmr.2007.07.004

*Toker, Y., & Ackerman, P. L. (2012). Utilizing occupational complexity levels in vocational interest assessments: Assessing interests for STEM areas. *Journal of Vocational Behavior, 80*(2), 524-544. doi:10.1016/j.jvb.2011.09.001

Top Employers Institute. (2013, March 1). Retrieved from http://www.top-employers.com/

Tracey, J. B., Tannenbaum, S. I., & Kavanagh, M. J. (1995). Applying trained skills on the job: The importance of the work environment. *Journal of Applied Psychology*, *80*(2), 239-252. doi:10.1037/0021-9010.80.2.239

Tracey, T. J. (1997). RANDALL: A Microsoft FORTRAN program for a randomization test of hypothesized order relations. *Educational and Psychological Measurement*, *57(*1), 164-168. doi:10.1177/0013164497057001012

Tracey, T. J., & Rounds, J. B. (1993). Evaluating Holland’s and Gati’s vocational-interest models: A structural meta-analysis. *Psychological Bulletin*, *113*(2), 229-246. doi:10.1037/0033-2909.113.2.229

Tranberg, M., Slane, S., & Ekeberg, S. E. (1993). The relation between interest congruence and satisfaction: A meta-analysis. *Journal of Vocational Behavior*, *42*(3), 253-264. doi:10.1006/jvbe.1993.1018

Trendence Institute. (2013, February 14). Retrieved from https://www.deutschlands100.de/deutschlands-100.html

Tsabari, O., Tziner, A., & Meir, E. I. (2005). Updated meta-analysis on the relationship between congruence and satisfaction. *Journal of Career Assessment*, *13*, 216-232. doi:10.1177/1069072704273165

Twitchell, S., Holton, E. F., & Trott, J. W. (2000). Technical training evaluation practices in the United States. *Performance Improvement Quarterly, 13*(3), 84-109. doi:10.1111/j.1937-8327.2000.tb00177.x

Tziner, A., Fisher, M., Senior, T., & Weisberg, J. (2007). Effects of trainee characteristics on training effectiveness. *International Journal of Selection and Assessment*, *15*(2), 167-174. doi:10.1111/j.1468-2389.2007.00378.x

Tziner, A., Haccoun, R. R., & Kadish, A. (1991). Personal and situational characteristics influencing the effectiveness of transfer of training improvement strategies. *Journal of Occupational Psychology*, *64*(2), 167-177. doi:10.1111/j.2044-8325.1991.tb00551.x

U.S. Department of Labor, Bureau of Labor Statistics. (2002). *Number of jobs held, labor market activity, and earnings* (Report No. (202) 691-7410). Retrieved from http://www.bls.gov/nls/nlsy79r19.pdf

United Nations Educational, Scientific and Cultural Organization (UNESCO). (2005). *Towards knowledge societies* [data file]. Retrieved from http://unesdoc.unesco.org/images/0014/001418/141843e.pdf

US eLearning Market Reaches $16.7 Billion in 2009. (2009, October 21). *Business Wire*. Retrieved from http://www.businesswire.com/news/home/20091021005329/en/eLearning-Market-Reaches-16.7-Billion-2009#.Uzk3QqIZnkd

Van Buren, M. E., & Erskine, W. (2002). *The 2002 ASTD state of the industry report.* Alexandria, VA: American Society of Training and Development Press.

Van de Vijver, F. R., & Leung, K. (1997). *Methods and data analysis for cross-cultural research*. Thousand Oaks, CA: Sage.

Van den Bossche, P., Segers, M., & Jansen, N. (2010). Transfer of training: The role of feedback in supportive social networks. *International Journal of Training and Development*, *14*(2), 81-94. doi:10.1111/j.1468-2419.2010.00343.x

*Van Iddekinge, C. H., Putka, D. J., & Campbell, J. P. (2011). Reconsidering vocational interests for personnel selection: The validity of an interest-based selection test in relation to job knowledge, job performance, and continuance intentions. *Journal of Applied Psychology, 96*(1), 13-33. doi:10.1037/a0021193

Van Iddekinge, C. H., Roth, P. L., Putka, D. J., & Lanivich, S. E. (2011). Are you interested? A meta-analysis of relations between vocational interests and employee performance and turnover. *Journal of Applied Psychology*, *96*(6), 1167-1194. doi:10.1037/a0024343

Van Knippenberg, D., & Schippers, M. C. (2007). Work group diversity. *Annual Review of Psychology*, *58*(1), 515-541. doi:10.1146/annurev.psych.58.110405.085546

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*(1), 4-70. doi:10.1177/109442810031002

Velada, R., Caetano, A., Michel, J. W., Lyons, B. D., & Kavanagh, M. J. (2007). The effects of training design, individual characteristics and work environment on transfer of training. *International Journal of Training and Development*, *11*(4), 282-294. doi:10.1111/j.1468-2419.2007.00286.x

*Vock, M., Köller, O., & Nagy, G. (2013). Vocational interests of intellectually gifted and highly achieving young adults. *British Journal of Educational Psychology, 83*(2), 305-328. doi:10.1111/j.2044-8279.2011.02063.x

Von Stumm, S., & Ackerman, P. L. (2013). Investment and intellect: A review and meta-analysis. *Psychological Bulletin, 139*(4)*,* 841-869. doi:10.1037/a0030746

Von Stumm, S., Hell, B., & Chamorro-Premuzic, T. (2011). The hungry mind: Intellectual curiosity is the third pillar of academic performance. *Perspectives on Psychological Science*, *6*(6), 574-588. doi:10.1177/1745691611421204

Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin, 117*(2)*,* 250-270. doi:10.1037/0033-2909.117.2.250

Wai, J., Cacchio, M., Putallaz, M., & Makel, M. C. (2010). Sex differences in the right tail of cognitive abilities: A 30 year examination. *Intelligence*, *38*(4), 412-423. doi:10.1016/j.intell.2010.04.006

Wakefield, J. A., & Doughtie, E. B. (1973). The geometric relationship between Holland's personality typology and the Vocational Preference Inventory. *Journal of Counseling Psychology*, *20*(6), 513-518. doi:10.1037/h0035196

Wall, J., Ahmed, V., & Smit, D. (2006). Addressing the lifelong learning needs of construction professionals using technology facilitated learning. *Journal for Education in the Built Environment*, *1*(2), 57-69. doi:10.11120/jebe.2006.01020057

Wang, G. G., & Wilcox, D. (2006). Training evaluation: Knowing more than is practiced. *Advances in Developing Human Resources*, *8*(4), 528-539. doi:10.1177/1523422306293007

Wang, W.-C. (2008). Assessment of differential item functioning. *Journal of Applied Measurement*, *9*(4), 387-408.

Wanous, J. P., Reichers, A. E., & Hudy, M. J. (1997). Overall job satisfaction: How good are single-item measures? *Journal of Applied Psychology*, *82*, 247-252. doi:10.1037/0021-9010.82.2.247

Webb, R. M., Lubinski, D., & Benbow, C. P. (2002). Mathematically facile adolescents with math-science aspirations: New perspectives on their educational and vocational development. *Journal of Educational Psychology, 94*(4), 785-794. doi:10.1037/0022-0663.94.4.785

Welsh, E. T., Wanberg, C. R., Brown, K. G., & Simmering, M. J. (2003). E-learning: Emerging uses, empirical results and future directions. *International Journal of Training and Development*, *7*(4), 245-258. doi:10.1046/j.1360-3736.2003.00184.x

Wenger, E. (1998). *Communities of practice: Learning, meaning and identity*. New York, NY: Cambridge University Press.

Wenger, E., McDermott, R., & Snyder, W. M. (2002). *Cultivating communities of practice*. Boston, MA: Harvard Business School Press.

Wetzel, E., & Hell, B. (2013). Gender-related differential item functioning in vocational interest measurement: An analysis of the AIST-R. *Journal of Individual Differences*, *34*(3), 170-183. doi:10.1027/1614-0001/a000112

Wetzel, E., Hell, B., & Pässler, K. (2012). Comparison of different test construction strategies in the development of a gender fair interest inventory using verbs. *Journal of Career Assessment*, *20*(1), 88-104. doi:10.1177/1069072711417166

Wilson, D. B. (2005). *METAREG Macro for SPSS* [computer software]. Available from http://mason.gmu.edu/~dwilsonb/ma.html

Wirth, R., Kauffeld, S., Bates, R., & Holton, E. F. (2009). Katalysatoren und Barrieren für den Transfererfolg: Das Lerntransfer-System-Inventar [Catalysts and barriers to success of transfer: The Learning Transfer System Inventory]. In S. Kauffeld, S. Grote, & E. Frieling (Eds.), *Handbuch Kompetenzentwicklung* (pp. 79-104). Stuttgart, Germany: Schäffer-Poeschel.

Wu, M. L., Adams, R. J., Wilson, M. R., & Haldane, S. (2007). *ConQuest* (version 2.0). Camberwell, Australia: Australian Council for Educational Research.

XING. (2013, January 2). *Forsa: Jeder dritte deutsche Arbeitnehmer ist bereit, im Jahre 2013 den Job zu wechseln.* Retrieved from https://corporate.xing.com/no_cache/deutsch/presse/pressemitteilungen/pressemitteilungen-detailansicht/article/studiebrforsa-jeder-dritte-deutsche-arbeitnehm/

Yamkovenko, B. V., Holton, E. F., & Bates, R. A. (2007). The Learning Transfer System Inventory (LTSI) in Ukraine: The cross-cultural validation of the instrument. *Journal of European Industrial Training*, *31*(5), 377-401. doi:10.1108/03090590710756819

Zell, E., & Krizan, Z. (2014). Do people have insight into their abilities? A metasynthesis. *Perspectives on Psychological Science, 9*(2), 111-125. doi:10.1177/1745691613518075

Zhao, Y., Lei, J., Yan, B., Lai, C., & Tan, S. (2005). What makes the difference? A practical analysis of research on the effectiveness of distance education. *Teachers College Record*, *107*(8), 1836-1884. doi:10.1111/j.1467-9620.2005.00544.x

Zieky, M. (1993). Practical questions in the use of DIF statistics in test development. In P. W. Holland, H. Wainer, P. W. Holland, & H. Wainer (Eds.), *Differential item functioning*. (pp. 337-347). Hillsdale, NJ: Erlbaum.

Appendix Tables

Appendix A

Mean Effect Size Estimates and Confidence Intervals for the Correlations between Holland's RIASEC Types and General Intelligence by Sex

	k	*N*	*r*	ρ	σ^2_ρ	% VE	90% CV	95% CI
Males								
Realistic	9	22,082	-.06	-.07	.015	3.3	[-.16, .09]	[-.15, .01]
Investigative	9	22,080	.36	**.42**	.016	2.8	[.26, .58]	[.33, .50]
Artistic	9	22,080	.13	**.15**	.006	8.6	[.05, .25]	[.10, .20]
Social	9	22,078	.10	.12	.020	2.6	[-.06, .30]	[.02, .21]
Enterprising	9	22,078	.02	.03	.006	8.5	[-.07, .13]	[-.03, .08]
Conventional	9	22,079	.02	.03	.003	17.2	[-.04, .10]	[-.01, .06]
Females								
Realistic	9	27,626	.13	**.15**	.002	16.2	[.09, .21]	[.11, .18]
Investigative	10	27,769	.28	**.32**	.003	11.4	[.25, .39]	[.28, .36]
Artistic	9	27,622	.22	**.25**	.018	2.3	[.08, .42]	[.16, .34]
Social	9	27,621	.15	.17	.036	1.2	[-.07, .41]	[.05, .30]
Enterprising	9	27,626	.04	.04	.003	14.0	[-.03, .11]	[.00, .08]
Conventional	9	27,624	-.07	-.08	.008	5.4	[-.20, .04]	[-.14, -.02]

Note. *k* = number of independent samples; *N* = total sample size; *r* = sample size weighted mean correlation; ρ = estimated true score correlation (corrected for sample error and unreliability); $\sigma^2\rho$ = estimated variance for true score correlation; % VE = percentage of variance in ρ accounted for by statistical artefacts; 90% CV = lower and upper bound of the 90% credibility interval for true score correlation; 95% CI = lower and upper bound of 95% confidence interval. Correlations are presented in boldface if the 90% credibility interval excludes zero.

Appendix B

Summary of Studies and Samples Included in the Meta-analysis

			Total							Males							Females						
ID	Author(s)[a]/Article		G	LD	I	QR	V	PS	MK	G	LD	I	QR	V	PS	MK	G	LD	I	QR	V	PS	MK
1	Ackerman (2000)	R		.09	.02						.02	-.08						.11	.05				
		I		.28	.14						.30	.06						.28	.18				
		A		.14	-.08						.13	-.06						.17	-.07				
		S		-.05	-.16						-.09	-.31						-.02	-.05				
		E		-.19	-.26						-.30	-.40						-.13	-.18				
		C		-.14	-.08						-.20	-.24						-.11	.02				
2	Ackerman et al. (2001)	R	.08	.02	.12																		
		I	.17	.21	.09																		
		A	.10	.20	-.02																		
		S																					
		E	-.16	-.11	-.16																		
		C	-.15	-.20	-.07																		
3	Ackerman et al. (1995)*	R		.14		.38	.24	.35		.18	-.05		.29	.18	.24		.36	.26		.37	.24	.44	
		I		.33		.34	.13	.31		.31	.31		.29	.13	.26		.30	.34		.31	.07	.32	
		A		.37		-.20	.01	-.08		.00	.34		-.36	.02	-.24		.30	.46		.13	.11	.24	
		S		-.08		-.14	-.04	.02		-.10	-.08		-.16	-.01	.02		.05	-.05		.12	.05	.17	
		E		-.32		-.15	-.15	-.06		-.36	-.51		-.17	-.16	-.04		-.16	-.21		-.06	-.11	-.03	
		C		-.32		.18	-.01	.14		-10	-.46		.24	-.01	.21		-.07	-.24		.12	-.02	.06	
4	Bergmann (2013)**	R	.20	.01		.18	.23			.04	-.10		.03	.16			.17	.04		.13	.19		
		I	.25	.12		.19	.21			.17	.09		.10	.17			.19	.11		.13	.16		
		A	.00	.17		-.14	-.03			.08	.22		-.07	.00			.10	.20		-.06	.05		
		S	-.18	-.06		-.16	-.16			-.09	.00		-.10	-.08			-.13	-.06		-.08	-.14		
		E	-.09	.00		-.08	-.10			-.08	.01		-.08	-.10			-.05	.02		-.04	-.07		
		C	-.02	-.01		.03	-.05			-.03	-.02		.02	-.07			.04	.01		.08	.00		
5	Carless (1999a)	R								-.09	-.07		-.09				.10	.03		.16			

	I							.15	.18		.10		.18	.18		.16	
	A							.03	.07		-.02		-.10	-.01		-.13	
	S							-.11	-.12		-.07		-.04	-.02		-.05	
	E							.07	.05		.09		.04	.06		.01	
	C							-.08	-.13		-.02		-.03	-.14		.02	
6 Carless (1999b)	R							.07		-.10		.01	-.12		.11		-.08
	I							.40		.35		.33	.33		.29		.22
	A							.05		-.04		.14	.24		.21		.16
	S							.10		.01		.17	-.18		-.17		-.10
	E							-.07		-.15		.05	-.05		-.05		-.02
	C							.18		.17		.15	-.15		-.29		.06
7 Carson (1996)	R	.26	.21	.09	.30	.02	.39										
	I	.40	.39	.27	.33	.16	.43										
	A	.18	.19	.07	.06	.08	.11										
	S	.03	.08	-.04	.02	-.01	-.06										
	E	-.23	-.16	-.26	-.17	-.13	-.11										
	C	-.14	-.19	-.04	-.18	.01	-.17										
8 Carson (1998a)	R	.15															
	I	.29															
	A	.21															
	S	.01															
	E	-.07															
	C	.12															
9 Carson (1998b)	R	-.09															
	I	.49															
	A	.13															
	S	.10															
	E	.13															
	C	.09															
10 Fritzsche et al. (1999)	R	.02															

	I	.11																						
	A	.07																						
	S	.06																						
	E	.09																						
	C	-.02																						
11 Kanfer & Ackerman (1996)	R		-.12		.05	.27	.03																	
	I		-.05		.11	.17	.02																	
	A		.19		.01	.16	-.03																	
	S		-.10		-.11	-.01	-.10																	
	E		-.30		-.17	-.17	-.02																	
	C		-.22		-.06	-.12	.16																	
12 Kaub et al. (2012)	R	.14	-.02			.35	.01			.12	-.10			.39	.04			.14	-.15			.27	.09	
	I	.06	.03			.23	-.05			-.03	-.07			.14	-.06			.10	-.01			.24	.00	
	A	-.15	.06			-.30	-.17			-.02	.25			-.21	-.20			-.22	.08			-.32	-.24	
	S	-.12	-.06			-.19	-.10			-.20	-.07			-.21	-.19			-.05	.05			-.12	-.11	
	E	-.04	.03			-.11	-.07			-.08	.10			-.11	-.20			-.01	.02			-.08	-.01	
	C	.06	-.05			.07	.02			.02	-.11			.04	-.09			.08	-.05			.07	.10	
13 Kelso et al. (1977)***	R																				-.01	.11	.04	.14
	I																				.27	.35	.02	.29
	A																				.00	.14	-.13	.02
	S																				-.03	-.30	-.04	-.15
	E																				-.03	-.12	.06	-.08
	C																				-.09	-.12	.15	-.07
14 Kirchler (1990)	R	-.19																						
	I	.29																						
	A	.22																						
15 Krapic et al. (2008)	R			.25																				
	I			.25																				
	A			.22																				
	S			-.05																				

	E			.02																			
	C			.18																			
16 Lowman et al. (1985)	I																	.28					
	C																						.14
17 Marcus et al. (2009)	R	.08																					
	I	.12																					
	A	.07																					
	S	-.07																					
	E	.05																					
	C	.07																					
18 Mussel (2013)	R		.05	.08							-.08	-.12							.14	.14			
	I		.24	.28							.26	.27							.25	.26			
	A		.28	.17							.23	.14							.32	.26			
	S		.06	-.01							.01	.05							.09	.03			
	E		-.02	-.01							.00	.01							-.03	-.03			
	C		.00	.12							.04	.18							-.02	.08			
19 Pässler & Hell (2012)	R	.31	.12		.39	.39				.24	.08		.26	.34				.26	.08		.27	.30	
	I	.31	.28		.28	.26				.27	.27		.22	.20				.26	.26		.18	.20	
	A	-.04	.12		-.24	-.13				.07	.20		-.07	-.09				-.01	.15		-.18	-.04	
	S	-.18	-.06		-.27	-.24				-.09	.01		-.11	-.21				-.14	-.05		-.19	-.15	
	E	-.09	-.14		-.02	-.11				-.11	-.14		-.10	-.17				-.11	-.16		-.05	-.12	
	C	.05	.00		.10	.02				-.04	-.04		.00	-.08				.08	.01		.12	.06	
20 Proyer (2006)	R	.22	.06		.15	.40				.28	.33		.00	.44				.16	-.08		.16	.37	
	I	.20	.10		.16	.25				.37	.34		.19	.49				.09	-.01		.10	.14	
	A	-.05	-.04		-.06	-.07				-.23	-.24		.01	-.42				.13	.09		.04	.12	
	S	-.26	-.26		-.26	-.21				-.31	-.35		-.10	-.34				-.18	-.20		-.11	-.13	
	E	-.04	-.13		-.13	-.02				-.10	-.22		.07	.01				-.03	-.10		.01	-.05	
	C	-.07	-.16		-.16	.05				.22	.03		.18	.39				-.20	-.23		-.14	-.08	
21 Randahl (1991)	R		-.03		.09	.34	-.13																
	I		.22		.23	.27	.08																

	A	.28		-.02	-.01	.13																	
	S	-.01		.01	-.13	.06																	
	E	-.17		-.04	-.06	-.04																	
	C	-.08		.15	.06	.04																	
22 Reeve & Heggestad (2004)	R								-.12								.11						
	I								.43								.31						
	A								.17								.28						
	S								.18								.25						
	E								.06								.06						
	C								.05								-.11						
23 Rolfhus & Ackerman (1996)	R	.23		.27	.28	.10	.35																
	I	.20		.14	.16	.02	.21																
	A	.24		.11	.06	.05	.11																
	S	.03		-.20	-.08	.14	-.21																
	E	-.03		.01	-.16	.02	-.15																
	C	-.05		.19	.01	.15	.04																
24 Schmidt et al. (1998)	R		.09		.20		.23																
	I		.14		.11		.08																
	A		.06		-.17		-.20																
	S		.00		-.21		-.28																
	E		-.04		-.07		-.11																
	C		.08		.00		-.06																
25 Stanley et al. (1995)	R												.13		.16						-.09		.11
	I												.08		.13						.11		-.08
	A												-.01		.00						.04		-.01
	S												-.20		-.06						-.01		-.08
	E												-.07		.03						-.09		-.08
	C												.06		.12						-.03		.11
26 Toker & Ackerman (2012[a])	R	.00		-.02	.21					.04		.00	.15					-.03		-.14	.15		
	I	.09		-.03	.11					.16		-.03	.14					.04		-.04	.14		

	A		.13	-.16	-.03		.05	-.18	-.01		.19	-.12	-.01
	S		.06	-.07	-.10		.00	-.05	-.06		.12	-.07	-.06
	E		-.10	-.11	-.23		-.20	-.05	-.23		-.03	-.16	-.23
	C		-.05	.15	.00		-.10	.18	-.10		-.02	.13	-.10
27 Toker & Ackerman (2012[b])	R		-.03	.12	.25		-.01	.11	.12		.04	.00	.30
	I		.10	-.05	-.01		.23	.12	-.06		.01	-.20	.00
	A		.16	-.19	-.10		.09	-.21	-.05		.34	-.14	-.10
	S		-.05	-.14	-.22		-.01	-.05	-.21		-.03	-.17	-.17
	E		-.16	-.05	-.27		-.08	.05	-.38		-.22	-.16	-.17
	C		-.21	.13	-.03		-.12	.11	-.19		-.26	.15	.15
28 Van Iddeking et al. (2011)	R	-.11											
	I	.07											
	A	-.09											
	S	-.12											
	E	-.08											
	C	-.15											
29 Vock et al. (2013)	R	.22				.15				.13			
	I	.24				.18				.18			
	A	-.09				-.04				-.02			
	S	-.19				-.11				-.13			
	E	-.04				-.10				.01			
	C	.03				-.08				.09			

Note. [a] Complete references can be found in the reference section. Cognitive abilities: G = General Intelligence. LD = Language Development. I = Induction. QR = Quantitative Reasoning. V = Visualization. PS = Perceptual Speed. MK = Mechanical Knowledge.

* PS complex is reported, PS simplex is also included in the analyses (Total: .08, .12, -.04, .07, -.05, -.07; Males: .04, .17, -.18, -.09, -.05, .08; Females: .15, .11, .10, .23, -.08, -.20). ** unpublished data. *** For MK Mechanical Comprehension is reported. Tool Knowledge, Space Perception, Shop Information, Automotive Information, Electronic Information are also included in the analyses.