

Tilburg University

Pigeonholing or learning instrument?

Lundgren, Henriette

Publication date:
2019

Document Version
Publisher's PDF, also known as Version of record

[Link to publication in Tilburg University Research Portal](#)

Citation for published version (APA):

Lundgren, H. (2019). *Pigeonholing or learning instrument? On the practice and perception of personality testing in human resource development*. Ridderprint.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

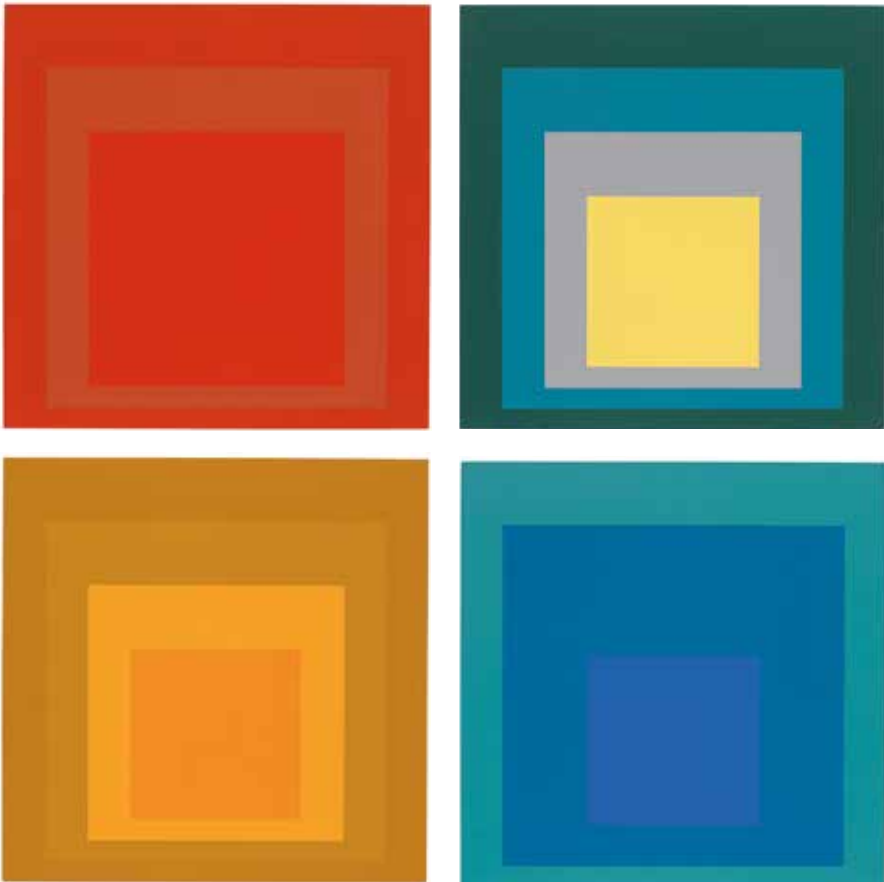
- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

PIGEONHOLING OR LEARNING INSTRUMENT?

On the Practice and Perception of Personality
Testing in Human Resource Development



Henriette Lundgren

PIGEONHOLING OR LEARNING INSTRUMENT?

On the Practice and Perception of Personality
Testing in Human Resource Development

Henriette Lundgren

Pigeonholing or Learning Instrument? On the Practice and Perception of
Personality Testing in Human Resource Development

Tilburg University

ISBN/EAN: 978-94-6375-337-1
Layout & Design by: Marilou Maes | persoonlijkproefschrift.nl
Printing: Ridderprint BV | www.ridderprint.nl
Cover images: Josef Albers
SP V, VI, XI, XII 1967
from the portfolio SP
Screenprint
Sheet: 24 1/4 x 24 1/4 in. (61.6 x 61.6 cm)
The Josef and Anni Albers Foundation, 1976.4.175.5,
1976.4.175.6, 1976.4.175.11, 1976.4.175.12
© 2019 The Josef and Anni Albers Foundation/Artists
Rights Society (ARS), New York/VG Bild-Kunst, Bonn
Photo: Tim Nighswander/Imaging4Art

Order via: henriette.lundgren@gmail.com

Copyright © 2019 Henriette Lundgren

All rights reserved. No part of this work may be reproduced, stored or transmitted in any way or by any means without the prior permission of the author, or when applicable, of the publishers of the scientific papers.

Pigeonholing or Learning Instrument?

On the Practice and Perception of Personality Testing
in Human Resource Development

Proefschrift

ter verkrijging van de graad van doctor aan Tilburg University
op gezag van de rector magnificus, prof. dr. E.H.L. Aarts,
in het openbaar te verdedigen ten overstaan van
een door het college voor promoties aangewezen commissie
in de Portrettenzaal van de Universiteit
op vrijdag 10 mei 2019 om 13:30 uur

door

Henriette Lundgren
geboren te Keulen, Duitsland

Promotiecommissie

Promotores: Prof. dr. R.F. Poell

Prof. dr. V.J. Marsick

Copromotor: Dr. B. Kroon

Overige leden van de Promotiecommissie:

Prof. dr. D. Gijbels

Prof. dr. M. Göhlich

Prof. dr. J.W.M. Kessels

Prof. dr. R.L. Martens

There are two kinds of people in the world:
people who think there are two kinds of
people in the world and people who don't.

Louis Menand,
What Personality Tests Really Deliver (2018)

Acknowledgements

It would not have been possible to produce this dissertation without the support of my family, friends, colleagues and many other contributors.

The cooperation with my Ph.D. advisors Rob Poell, Brigitte Kroon, and Victoria Marsick has been most inspiring and fruitful. I have great appreciation for your continuous support and belief in me. My thanks go to Rob for traveling across the Atlantic for research meetings and conferences, which made the remote supervision more tangible and productive. Special thanks to Victoria's husband Peter Neaman who lovingly pushed me to continue my dissertation when my initial arrangement fell through.

Many practitioners provided access to the field and support throughout my research. I would like to thank all participants of my studies: human resource development practitioners, industry experts, and test takers. This dissertation depended on your insights, comments, and reflections. Thanks also to numerous anonymous librarians at the Bodleian, Bobst, Mann, and Olin Libraries for your guidance, patience and reference support.

My University of Hamburg colleagues Eva-Christine Kubsch, Barbara Nienkemper, and Wibke Riekmann helped me to get started during the initial phase of this project. Special appreciation goes to Barbara Kohlstock, Rebecca Lingwood, Jonathan Michie, Anita Pachner, Matthias Rohs, Gabriela Sewz, Clare Wakeham, Marianne van Woerkom, and Rita Walczuch for their support and mentorship at various stages of my academic journey.

I am grateful to my Teachers College research collaborators and co-authors, from whom I draw inspiration and guidance, including April Bang, Molly Clark, Sean Justice, Jennifer Yates, Lyle Yorks, and SeoYoon Sung. Thank you for the many years of writing and discussion... or shall we reflect a bit more on that?

I have greatly enjoyed the intellectual exchanges with Werner Deetz, Steven Nishida, Allan Renz and Ranjit Singh. Thanks, Ranjit, for helping me develop my teaching and research statements.

Friendship means a lot to me. I therefore would like to thank Katja Lubina, my best friend forever, acrobatics buddy, witness to my wedding, and most creative person I know, who supported me relentlessly from start to finish of this project. I also want to say thank you to my Maastricht Mädels Diana Schramm, Annette Zimmermann, and Vanessa Lemarié for their encouragement. To Nynke Schilder Hosseinion for her lasting friendship and hospitality during my visits to Breda and Tilburg. To my Oxford friends: Lucie Abeler-Dörner, Luce Bellefontaine, Ulrike Bilgram Przepiórka, Jacqueline Denzer, Maxi Freund, Heidrun Jöken, Anna Kjellin, Kathrin Lebrecht, Chikako Mita, Martina Wittmann-Hohlbein, and Susanne Spiegel who all know what's it like to live with at least one foot in academia. Thanks for your support during our time as Newcomers and beyond. To my Ithaca friends: Sheila Stone and Charlie Strohman, Julia Franke, Christiane Schnabel, Raf Chew, Jing-Ting Huang, Danielle Fuld, Katja Nowack, and Iwijn de Vlaminck for their culinary care, proofreading support, curiosity, and compassion. To my friends back in Europe: Ferhat Akkaya, Anna Beck, Simone Doerfner, Corina Krapf, Kornelia Matthes, Susanne Pudwell, Joe and Anne Trapp, and Ineke Verhagen, who cheered for me from afar.

I would also like to thank my colleagues at Unilever (who introduced me to personality testing), Grange Partnership, Byrne Dairy, and now Corning who have given me professional advice and kind guidance throughout my career.

Finally, I would like to thank my brother Anderš, my parents Nils and Friederike, and Tante & Onkel Ressel for their unconditional love over the last four decades. And of course, thank you, Malte. Without you, life (and this project) would have been half as much fun. Seriously.

Henriette Lundgren
Ithaca, NY in March 2019

Table of contents

Acknowledgements	8
List of tables	12
List of figures	12
List of abbreviations	13
Chapter 1. General introduction	15
Definitions	18
Intended contributions	22
Research outline	23
Overview of chapters	25
Chapter 2. Personality testing and workplace training	29
Abstract	30
Introduction	31
Theoretical background	32
Research questions	34
Methodology	36
Findings – country context	38
Findings – emergent themes	44
Conclusions and discussion	55
Chapter 3. On critical reflection	63
Abstract	64
Introduction	65
Methodology: Conducting a step-by-step literature review	70
Approaches to operationalize critical reflection	73
Assessing levels and reflection outcomes	80
Four ways of improving critical reflection research	85
Chapter 4. “This is not a test”	93
Abstract	94
Introduction	95
Theoretical foundation	98
Methods	103
Strategies in HRD test use	110
Discussion	120
Conclusion	126

Chapter 5. Pigeonholing or learning instrument?	129
Abstract	130
Introduction	131
Theory: Test takers' reactions	134
Methods	138
Findings – factors that explain test takers' reactions	143
Findings – emerging influencing factors to test takers' reactions	148
Discussion	150
Chapter 6. Discussion and Conclusion	157
Main findings	158
Overarching contribution	163
Reflections on theories	164
Reflections on methods	166
Implications for future research	169
Implications for practice	171
Conclusion: Pigeonholing and learning instrument	172
Summary	175
Zusammenfassung	187
References	199
Appendices	223
Appendix 1: Personality test names and abbreviations	224
Appendix 2: Data collection overview of multiple-case study	225
Appendix 3: Interview guide practitioner study	228
About the author	231

List of tables

Table 2.1 Associations and testing centers	40
Table 2.2 Popular personality tests in Germany, the UK and the Netherlands	42
Table 2.3 Overview of test publishers with challenges	45
Table 2.4 Purposes for using personality tests	51
Table 2.5 Comparison of NEO PI-R and MBTI marketing leaflets	54
Table 3.1 Summary of approaches to operationalize critical reflection	74
Table 3.2 Summary of assessing levels and reflection outcomes	78
Table 4.1 Overview of HRD participants	106
Table 4.2 Strategies in personality test use	109
Table 5.1 Hypothesized factors that influence test takers' reactions	135
Table 5.2 Overview of test taker participants	140
Table 5.3 Units of analysis in QCA	141
Table 5.4 Results of cross-case comparison	144

List of figures

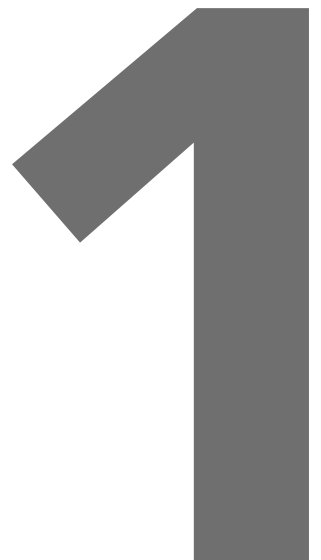
<i>Figure 1.1</i> Internal and external stakeholders	21
<i>Figure 1.2</i> Dissertation outline by chapter	23
<i>Figure 2.1</i> Data collection overview	38
<i>Figure 2.2</i> Stakeholders in personality testing industry	44
<i>Figure 3.1</i> Timeline and number of citations of Mezirow's critical reflection work	66
<i>Figure 3.2</i> Search terms used for literature review	71
<i>Figure 3.3</i> Literature review process steps and outcome	73
<i>Figure 5.1</i> Longitudinal design in three phases	138
<i>Figure 5.2</i> Factors that influence test takers' reaction to personality testing	150

List of abbreviations

AC	Assessment Center
BDP	Professional Association of German Psychologists
BPS	British Psychological Society
CIPD	Chartered Institute of Personnel and Development
COTAN	Dutch Committee on Testing Affairs
HR	Human Resources
HRD	Human Resource Development
HRM	Human Resource Management
EFPA	European Federation of Psychological Associations
L&D	Learning and Development
MD	Management Development
NA	Not applicable / Not available
NIP	Dutch Institute of Psychologists
NVP	Dutch Association for Personnel Management and Organization Development
PTC	Psychometric Testing Center
TBS-TK	Test Review System of the Board of Assessment and Testing
T&D	Training and Development
QCA	Qualitative Content Analysis

Chapter 1

General introduction



“I was typed a blue type and ... my manager really stereotyped me. By this I mean that he told me several times in meetings or face-to-face that I should try to be more of a different color because I would make the working situation more difficult or less pleasant. He was bright yellow and wanted people to try to be as positive and less organized.” (Interview quote by Sophie, 395-399; Education specialist who decided not to apply personality testing in developmental settings after this experience as a test taker.)

“Sometimes, talking about the reliability and the validity behind the tool is useful. Other times, it’s about perhaps being a bit softer in terms of your approach and saying, ‘you know, the purpose of it isn’t to try to put people into boxes. It’s to try to help you to understand yourself better.’” (Interview quote by Jackie, 398-401; Learning and development manager who explains what is important when introducing a personality test in management development.)

The two quotes above stem from Human Resource Development (HRD) professionals who reflect on their experience with personality testing tools in the workplace. In the practice of HRD, tools are ubiquitous. They include devices, systems, methodologies and approaches that are intended to support the development of an individual employee or a group of employees. For example, action learning can be seen as a tool to make workplace learning more problem-based (O’Neil & Marsick, 2007); mobile devices can be used as tools to support micro- and self-paced learning (ATD, 2017); and personality tests can be employed as tools for self-discovery in leadership development (Allen & Hartman, 2008).

While the structure and concepts of personality can be outlined along different schools of thought (cf. Friedman & Schustack, 2010), personality can be defined as the “psychological processes that determine a person’s characteristic behavior and thought” (Allport, 1961, p. 28). A personality test then describes a standardized assessment used to determine a person’s set of psychological processes, preferences, traits or behavioral styles. Most commonly, but not exclusively, a self-report questionnaire is used where the participant of such an assessment – the test taker – answers a series of questions. A test report is created from the answers of the questionnaire and feedback is given to the test taker to point out areas for personal development. Many personality tests

were originally created to support the personnel selection and hiring decisions (Hossiep, Schecke, & Weiss, 2015). However, survey research conducted by the Chartered Institute of Personnel and Development (CIPD) in the UK shows that personality tests are increasingly employed by HRD professionals as a tool of their professional practice (McGurk & Belliveau, 2012).

However, while personality tests as tools in HRD practice are often promoted as enablers to enhance personal development and organizational functioning, they can also yield the opposite effect of departmentalization and stereotyping, as Sophie's quote at the beginning of this introduction shows. Personality tests can be criticized in many ways, and one of the contradictions evolves around the static nature of personality itself. While there is disagreement among personality theorists regarding the plasticity of personality traits (McCrae & Costa, 2003; Roberts & Mroczek, 2008; Srivastava, John, Gosling, & Potter, 2003), popular personality test literature and test manuals state that personality develops until early adulthood and then remains relatively stable over the course of one's life (Keirsey & Bates, 1984; Myers, 1962; Williams, 2012). Assuming that personality is a static construct, it leaves scanty room for change and development. However, HRD at its core is about seeking opportunities for learning, growth and development. The question thus arises how a test that determines "what is" can be effectively used in an adult education setting that strives to find out "what can be". Do HRD professionals recognize this paradoxical situation where a static construct of personality is used in a formative context? If yes, how do HRD professionals reflect on their practice, making meaning and/or sense of it? The quote by Jackie suggests that reflecting upon the practice and perception of personality testing in HRD is necessary to deliver training and development in an ethical way.

The overall research question for this dissertation is: How are personality tests used in HRD? In order to understand this, the research explores the practice of personality testing in settings such as management development. The research also investigates the effect that this practice has on participants' perceptions of personality testing practice. Special attention is given to the role of critical reflection as an indicator for deliberate professional practice with regard to the purpose of HRD professionals and test takers. In summary, the research present-

ed in this dissertation aims to explore the practice and perception of personality testing in HRD in order to build on ethical and effective practice in the field.

Definitions

Personality

According to the trait school of thought, as represented by Gordon Allport (1897-1967), personality can be summarized as a person's characteristic behaviors and thoughts, measured along a unique set of dimensions. Allport, a leading 20th century personality researcher and advocate of the trait approach, emphasizes the basic dimensions of personality that form a person's unique styles and dispositions. However, other personality theorists, such as Carl Jung (1875-1961), a prominent personality theorist who lived during the same time as Allport but who represents the neo-analytic approach of personality theory, emphasizes the self as it struggles to deal with emotions and drives on the inside that respond to the demands of others from the outside. In his writings, Jung refers to the different categories of responses as "archetypes" or "psychological types" (Jung, 1923). While these two approaches – Allport's trait and Jung's type theory – are largely different, the vast majority of personality tests available on the market today can be characterized as either trait-based test or type-based tests. It should be noted that this brief overview on trait and type personality theory leaves out other approaches, such as behaviorism or interactionism. However, this distinction between trait and type that is often referred to in personality literature and research is sufficient to introduce the reader of this dissertation to underlying aspects and assumptions that form the basis of commonly available personality tests.

Personality tests

Looking further at the distinction between type and trait based psychometrics, personality tests differ along a number of dimensions, including their input and output formats. *Type-based* tests use a limited number of clearly distinct, non-overlapping personality types to describe people. In comparison, *trait-based* tests aim to describe preference dispositions that can vary in strength and that can be compared with those of other people in a norm group (Gerrig & Zimbardo, 2009). Psychometric test qualities help practitioners to distinguish among the many personality tests available on the market. The two test qualities that prac-

tioners refer to most often are reliability and validity. Reliability describes the extent to which the test measures personality consistently while test validity indicates the extent to which the test in fact measures that facet of personality that it intends to measure.

Another differentiator among tests is the test scoring system used. Trait-based tests usually make use of Likert scales and norm-referenced scoring where the test taker's result is compared with results of other people of a certain population, e.g., managers. In comparison, type-based tests more commonly work with ipsative scoring formats, in which test takers are asked to choose between two alternatives and often opposing poles (Salgado & Tauriz, 2014).

Personality testing industry

While the number of distinct personality theories and empirically derived personality models is small, a large number of about 2,500 personality tests exist in the market, with a few million tests taken annually (Moore, 1987). At the beginning of this dissertation research, it was estimated that about \$400m in test revenue is generated every year (Talbot, 1999). This number has since been increased to \$500m (Weber & Dwoskin, 2014) and was most recently estimated to have risen to \$2bn in revenue (Emre, 2018; E. Gray, 2015). A large proportion of these revenues are generated through HRD activities, where personality tests are used for individual and team development (McGurk & Belliveau, 2012).

Human resource development

HRD is a vast field that is mainly concerned with the transformational aspects of personnel management. Watkins (1989) defines HRD as “the field of study and practice responsible for the fostering of a long-term, work-related learning capacity” (p. 427). Capacity building can take place at different levels, including the individual, teams and entire departments and organizations (cf. Yorks, 2004). The emphasis here is on the learning capacity, that is, the formative aspect of practice.

The field of HRD practice can be divided into three areas, all of which use development as their primary process: career development (CD), organization development (OD), and training & development (T&D). The latter describes those activities within HRD that help identify, assess and develop key competencies

that support individuals and teams to best perform current or future jobs, often through formal learning. In T&D, an HRD professional creates a program or curriculum that the learner pursues by following the directions of the trainer through listening, small group discussions or work on capstone projects. Guided coaching outside the classroom often supports formal learning structures and its learners.

In the context of T&D, the term “transfer of training” plays an important role as it describes “the degree to which knowledge, skills, and abilities learned in training are applied to the job” (Bates, Cannonier, & Hatala, 2014, p. 386). HRD professionals evaluate their training outcome using transfer of training measures with the aim to help organizations decide whether a certain T&D program has been successful or not. Transfer of training also plays a role when it comes to evaluating outcomes of HRD programs where personality tests have been used.

Personality testing in HRD

The most widespread use of personality tests in the workplace is through T&D workshops, especially those that are dedicated to the development of managers and executives (Goodstein & Lanyon, 1999). Many different personality tests exist, with the type-based Myers-Briggs Type Indicator (MBTI) being a popular instrument in management development. By comparison, research studies make more use of trait-based personality assessments that are based on the Five Factor Model (or: Big Five), such as the NEO PI-R (see Appendix 1 for more information on test names and abbreviations).

The personality testing industry is unregulated. This means that there is no central governing body or authority that organizes how personality tests are created, tested, distributed and applied in practice. Concurrently, various internal – those employed by an organization – and external stakeholders – those outside an organization – have a vested interest in personality tests and how they are used in developmental settings (see Figure 1.1).

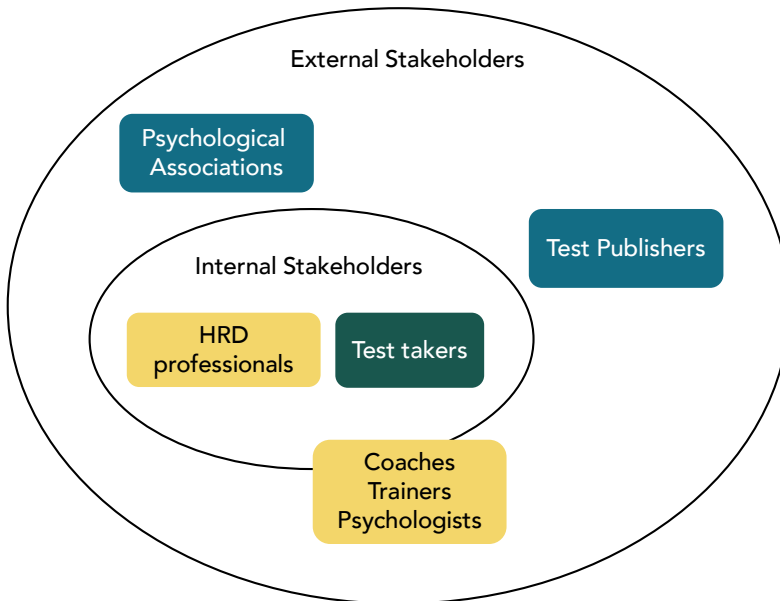


Figure 1.1 Internal and external stakeholders

For example, a psychological association is interested in ensuring that its members apply psychometric instruments in a standardized and ethical way (Ackerschott, 2000; Ackerschott, Gantner, & Schmitt, 2016). This interest potentially conflicts with the interests of test publishers who aim to increase the number of test licenses sold (Furnham & Jackson, 2011). Internal HRD professionals, on the other hand, hold an interest in creating inclusive work environments (Watkins, 1989) that contribute to the performance goals of the organization (R. A. Swanson & Holton, 2001), while coaches, trainers and psychologists often act on behalf of their client when choosing a test. Lastly, employees who participate in workplace training events as internal stakeholders have an interest in advancing their careers through personal development (Goodstein & Prien, 2006). The distinct interests can lead to challenges with regard to effective and ethical practice in HRD, where ethical practice refers to professional integrity, diligence and quality delivery (Sullivan, 2005) and effective practice to the increase of training transfer into the workplace (Gegenfurtner, Veermans, Festner, & Gruber, 2009).

Intended contributions

In this web of different stakeholders and interests, the question of purpose and perception arises: How are personality tests put into practice? Are they used as a learning instrument, and also perceived as such, or do test takers view them predominantly as a tool for pigeonholing, even if not desired by associations, publishers and HRD professionals? The intended contributions of this research therefore are to:

- Highlight industry dynamics in order to clarify what the role of professional associations is and could be.
- Develop a methodology based on existing empirical critical reflection research in order to identify practical ways how tool use in the workplace can be studied through reflection.
- Explore how experienced HRD professionals approach personality testing in their own practice with an expected result to give guidance to newer practitioners who are entering the field.
- Establish a framework of test takers' perceptions of personality testing in HRD in order to advance the literature in developmental contexts.
- Encourage critical reflection of and enhanced discourse on the ethics of test use for improved professional practice in HRD.

As much as this research could have been about specific personality tests and their psychometric qualities, a conscious decision was made not to focus on the tests themselves. Instead, attention is given to the everyday practice of HRD practitioners when it comes to test use, as well as the perception of test takers on this practice. It should be noted that this research is not the personality test of anyone who participated in this study: neither the practitioners nor the test takers who were interviewed and observed.

Research outline

The main aim of this research is to explore the practice and perception of personality testing in HRD with the goal of establishing ethical and effective practice in the field. Inspired by evidence-based management literatures (Briner, Denyer, & Rousseau, 2009; Reay, Berta, & Kohn, 2009), this research was conducted drawing from different sources of information, including critically reviewed research studies, evidence from the local context of practice, HRD practitioners' lived experience and judgment, and the views of test takers who are affected by the practice. The dissertation consists of four studies that are presented in Chapters 2, 3, 4 and 5 (see Figure 1.2).

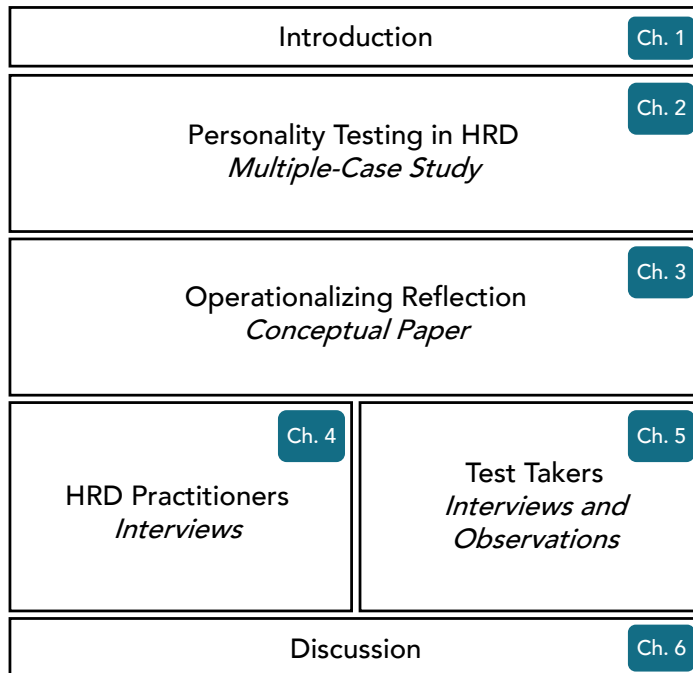


Figure 1.2 Dissertation outline by chapter

Objectives

The research is structured around four studies that each has distinct research objectives.

- The focus of the first study summarized in Chapter 2 is on the personality testing industry. This study inquires about the role of internal and external stakeholders, the value of psychometric and practical considerations in test selection, and the purpose of personality test use in workplace training.
- The focus of the second study – in the form of a literature review described in Chapter 3 – is on critical reflection. This study inquires about existing reflection research and how its operationalization informs the current and future research.
- The focus of the third study is on HRD professionals and their personality testing practice. Summarized in Chapter 4, this study illustrates individual meaning-making and organizational sensemaking structures of practitioners.
- The fourth study presented in Chapter 5 focuses on test takers who participate in management development workshops where a personality test is used. This study inquires about differences in test takers' reactions and factors that influence these.

Chapter 6 concludes this research with a discussion of its main findings, limitations and recommendations for future research.

Research context and methods

The research was conducted in Western Europe, with a focus on personality testing in the UK, Germany and the Netherlands. With the exception of a few international HRD practitioners from Denmark and the United States (Chapter 4), all study participants were based in one of these Western European countries. A set of qualitative methods was employed, combining multiple-case study, qualitative interviews, and participatory observations. Resulting data was analyzed using open coding, inductive analysis (Miles & Huberman, 1994), qualitative

comparative analysis (Rihoux & Ragin, 2009), and constant comparison as part of grounded theory (Glaser & Strauss, 1967).

Data access and data collection

As an HRD practitioner who has worked in workplace learning and related development fields since 2003, the author of this study has had personal experience with personality tests as well as professional access to tools, stakeholders and organizations in the testing industry. Through these existing professional connections as well as new connections made at industry conferences and trade-shows, the multiple-case study data was collected for Chapter 2 between 2012 and 2016. In parallel, the first set of test-taker interviews that contribute to Chapter 5 were recorded in mid 2012, based on a convenience sampling strategy. From here, access negotiations with other organizations were held resulting in a second set of participatory observations and interviews at the end of 2012 and beginning of 2013. For Chapter 4, a call for voluntary research participation was sent through the author's professional networks, and the data collection for the HRD practitioner study was concluded in December 2016. In order to ensure consistency in the interview process, the author personally collected all data that contributes to these three empirical studies. Because of the author's professional experience in HRD it is acknowledged that her positionality will have affected the research findings presented in this dissertation.

Overview of chapters

Chapter 2. How can the personality testing industry be described with its tools, stakeholders and dynamics?

The aim of the first study is to paint a “landscape picture” of the personality testing industry and how it relates to workplace training. This study explores why personality tests are used while inquiring about the role of different stakeholders, the value of psychometric and practical considerations in test selection, and the purpose of personality test use in a developmental context.

The following questions are addressed: 1. How do external stakeholders, e.g., publishers, associations and psychologists, engage with internal stakeholders, e.g., HR practitioners, when choosing a personality test for workplace training?;

2. How are practical and psychometric considerations weighted during the decision-making process? and 3. How are personality tests positioned by HRD practitioners for the purpose of workplace training? This chapter uses a multiple-case study approach where ethnographic data collected from publishers, associations, psychologists and HRD practitioners in Germany, the UK and the Netherlands is analyzed and compared.

Chapter 3. How has existing literature conceptualized and operationalized critical reflection?

The aim of the second study is to develop a methodology based on existing empirical research on critical reflection, in order to identify practical ways to study tool use in the workplace through reflection. This chapter reviews empirical studies from adult education, leadership development, management learning and higher education spectrum that each research critical reflection based on Jack Mezirow's (1978, 1981, 1990, 1991, 1998) definition. The aim is to come up with a framework of operationalization to guide future reflection research.

The following questions are addressed: 1. What approaches have been used to operationalize critical reflection?; and 2. How have levels of reflection been assessed, and with what outcomes? This study is conducted in the form of a literature review where 12 research studies on critical reflection are dissected, analyzed, and compared. This process results in the description of four suggestions for improvements.

Chapter 4. How do HRD professionals make sense of their own personality testing practice?

The aim of the third study is to explore the lived experiences of HRD professionals who engage with personality tests as tools in their professional practice. This "practitioner study" seeks to describe the personality test insights, attitudes and approaches in order to make sense of the paradoxical situation wherein personality tests are popular among HRD professionals but at the same time are one of the most critiqued instruments.

The following question is addressed in this study: How do individual meaning-making and organizational sensemaking theories help to explain the widespread and sustained use of personality tests among HRD professionals? This study uses a grounded theory approach when conducting and analyzing 18 qualitative interviews through inductive data analysis.

Chapter 5. How do test takers react to personality testing in developmental contexts?

The aim of the fourth and final study is to depict the lived experiences of workshop participants who have been confronted with personality testing as part of a management development program at work. This “test taker” study inquires about different aspects of the workshop setting, tools and facilitator in order to verify whether influencing factors from management selection are also relevant in the HRD context.

The following question is addressed in this study: Which factors can explain differences among test takers’ reactions to personality testing in the context of management development? This study employs a qualitative longitudinal approach with three phases of data collection, including two participatory observations and eleven semi-structured interviews with test takers. I analyzed the data using Qualitative Comparative Analysis (QCA).

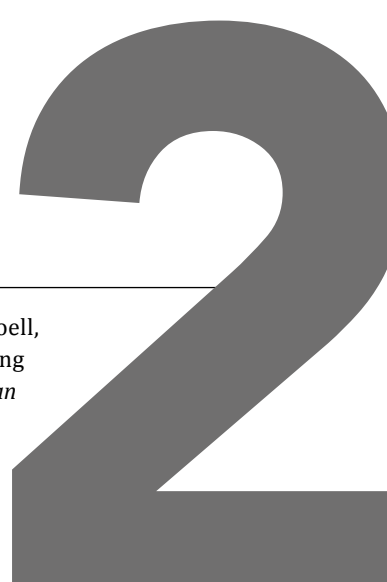
In Chapter 6, the final chapter of this dissertation, the main findings are discussed and conclusions are presented. It also contains reflections on theory and methods, followed by implications for future research and practice. The chapter closes with a discussion of this dissertation’s overarching question: “Pigeonholing or learning instrument?”.

Chapter 2

Personality testing and workplace training

Exploring stakeholders, products and purpose in Western Europe¹

¹ This chapter has been published as: Lundgren, H., Kroon, B., & Poell, R. F. (2017). Personality testing and workplace training: Exploring stakeholders, products and purpose in Western Europe. *European Journal of Training and Development*, 41(3), 198–221.



Abstract

Purpose: The purpose of this research paper is to explore how and why personality tests are used in workplace training. The research is guided by three research questions that inquire about the role of external and internal stakeholders, the value of psychometric and practical considerations in test selection, and the purpose of personality test use in workplace training.

Methods: This research paper employs multiple-case study analysis. Interviews, test reports, product flyers and email correspondence were collected and analyzed from publishers, associations, psychologists and HRD practitioners in Germany, the UK and the Netherlands between 2012 and 2016.

Findings: Themes emerge around industry tensions among practitioners and professional associations, psychologists and non-psychologists. Ease-of-use is a more important factor than psychometrics in the decision-making process. Also, practitioners welcome publishers that offer free coaching support. In the process of using tests for development rather than assessment, re-labeling takes place when practitioners and publishers employ positive terms for personality tests as tools for personal stocktaking and development.

Research limitations: Despite extensive data collection and analysis efforts, this study is limited by its focus on a relatively small number of country cases and stakeholders per case.

Practical implications: By combining scientific evidence with practical application, stakeholders can take first steps towards more evidence-based HRD practice around personality testing in workplace training.

Originality/value: Little academic literature exists on the use of personality testing in workplace training. Without a clear understanding of the use of personality testing outside personnel selection, the current practice of personality tests for developmental purposes could raise ethical concerns about the rights and responsibilities of test takers.

Introduction

Personality inventories like the Myers-Briggs Type Indicator (MBTI) or Big Five assessments have been used and marketed widely, and they enjoy great popularity in business, human resource development (HRD) and adult education settings. Psychometrics and personality testing are fascinating fields of study, with numerous writings on its history (Hough & Oswald, 2005; McAdams, 1997), its development (Hough & Oswald, 2008) and its role in the workplace (Barrick & Mount, 2005; Diekmann & König, 2015). Originally, personality tests were designed for personnel selection (DiMilia, 2004; Goodstein & Lanyon, 1999; Lievens, van Dam, & Anderson, 2002; Tett & Christiansen, 2007). Increasingly, they are also used in developmental areas of the human resources (HR) spectrum, such as coaching (Nelson & Hogan, 2009; Passmore, 2008; Passmore, Holloway, & Rawle-Cope, 2010; Scharlau, 2004), educational leadership (Tomlinson, 2004), organizational and team development (Badham, Garrety, Morigan, Zanko, & Dawson, 2003; Clinebell & Stecher, 2003; Ludeman, 1995), and management training (Ford & Harding, 2007; Furnham & Jackson, 2011; Goodstein & Prien, 2006).

It is estimated that four out of ten tests are used in team building and management development, often referred to as tools for self-exploration and self-reflection, and with the aim to improve team performance (Gardner & Martinko, 1996). Often, non-psychologists administer those tests (Bartram, 2001) with limited knowledge on the tests' psychometric properties (Furnham, 2008). While personality test use remains "a hot and continually debated topic" (Furnham, 2004, p. 26), very little literature exists that explains the use of personality tests in team development, coaching and workplace training. Earlier data suggests that the market size is impressive: about 2,500 tests are administered a few million times every year (Moore, 1987), producing a turnover of around \$500 million per annum (Paul, 2005; Talbot, 1999; Weber & Dwoskin, 2014). Since new technologies support the quick creation and dissemination of tests, it can be assumed that the personality test market is growing even further. This means that employees encounter with great certainty a moment in their career when their personality is assessed in organizational life. Without a clear understanding of the use of personality testing for purposes other than selection, the current practice of using personality tests for development could

raise ethical concerns about the rights and responsibilities of test takers (Smith & Smith, 2005). For example, test takers might be given the impression that they are being assessed, even if that is not the aim in a workplace training setting. Also, test takers might feel uncomfortable sharing test feedback at work with their line managers and colleagues.

This paper explores why and how personality tests are used in the practice of workplace training. A multiple-case study approach was chosen to explore country-specific differences and to find emerging patterns that depict the dynamics in this industry. First, existing empirical studies will be reviewed, and research questions will be formulated. Second, the study's methodology will be laid out, including the rationale behind selecting multiple cases. Next, findings will be reported concerning different industry stakeholders, the decision-making process of choosing one test over another, and the purpose of using tests in workplace training. The paper closes with a discussion stating research and practical implications for the HR community.

Although this paper started with an example of the MBTI instrument, it does not center on the MBTI or any other personality inventory in particular. Instead, the paper explores broadly the personality testing industry and its players, dynamics, and dilemmas².

Theoretical background

Defining personality

Personality psychology looks at the individual differences of people, including preferences, motives and predispositions. Comparable to different schools of thought in the social sciences, different perspectives on personality psychology exist, including psychoanalytic, behaviorist, trait and interactionist perspectives (Friedman & Schustack, 2010). Gordon Allport (1897-1967), a leading personality researcher of the 20th century and a proponent of the trait approach, defined personality as the “psychological processes that determine a person’s characteristic behavior and thought” (1961, p. 28). This broad definition is useful for this

² For readers interested in MBTI’s use in management, we recommend the following reviews and analyses: Gardner and Martinko, 1996; Kuipers et al., 2009; Michael, 2003; and Pittenger, 2005 as a starting point.

paper on personality testing in HRD, as it includes individual characteristics as well as motives and attitudes.

Assessing personality

Personality can be assessed using psychometric tests; these are standardized ways for determining a person's set of preferences, most commonly but not exclusively using self-report questionnaires. Other measurements of personality include projective and objective tests (Kline, 1976). Different personality tests may be referred to as *taxonomies*, *assessments*, *inventories* or *instruments*; however, practitioners and practitioner literature more commonly refer to them as "tests". For simplicity, this paper uses the blanket term "test" when writing about any self-report questionnaire.

The most widely distributed personality tests in HRD can be classified as either trait or type tests. While type tests use a limited number of clearly distinct, non-overlapping personality types to describe people, trait tests aim to describe preference dispositions that can vary in strength and that can be compared with those of other people in a norm group (Gerrig & Zimbardo, 2009). Psychometric test qualities help distinguish among the many personality tests available on the market; most notably test reliability, that is the extent to which it measures personality consistently, and test validity, that is the extent to which it measures that facet of personality that it intends to measure. Another differentiator between tests is the test scoring system employed. Trait taxonomies usually make use of Likert scales and norm-referenced scoring where the test taker's score is compared with the scores of other people. In comparison, type taxonomies more commonly work with ipsative scoring formats, in which test takers are asked to choose between two alternatives and often opposing poles (Salgado & Tauriz, 2014).

Assumptions and criticism

The US Army first made psychometric testing available in workplace settings for selection purposes, where there was a strong focus on ability. Personality instruments like the *MBTI* and *Big Five* were later designed and used by psychologists more to measure personality traits and less as a screening tool. Personality tests are based on certain assumptions, for example that personality remains relatively stable over time and across different situations, which is the reason why

personality tests were initially used for selection. If personality can be tested before hiring, and if personality preferences remain the same over time and in different contexts, then better hiring decisions could be made according to personality theory and research (Barrick & Mount, 1991; Barrick, Mount, & Judge, 2001; J. Hogan & Holland, 2003). Several authors have debated the assumption of trait stability over time (Lucas & Donnellan, 2011; Roberts & Mroczek, 2008; Srivastava et al., 2003) and the predictability of behavior using preference and leaving out situational influences (Mischel, 1968; Tett & Burnett, 2003), but that has not stopped practitioners from using personality tests in the workplace. Personality psychology approaches and trends develop over time, and so has the use of personality testing. Currently, the most common description of individual differences in academic research is the *Big Five* trait taxonomy (Furnham, Eracleous, & Chamorro-Premuzic, 2009; Larson, Rottinghaus, & Borgen, 2002; Wolff & Kim, 2012), which is based on the *Five Factor Model* of personality (Goldberg, 1990; McCrae & Costa, 1999; McCrae & McCrae, 1996).

Research questions

Where only a limited number of studies exist on the use of personality testing in workplace learning, a more holistic approach is needed to comprehend the practice and decision-making rules of those working in the field. Evidence-based management offers a model that integrates four sources of information (Briner et al., 2009; Reay et al., 2009): critically-reviewed research evidence is complemented with practitioner experience and judgment, evidence from the local context and the views of those stakeholders who are affected by the practice.

External and internal stakeholders

In organizational life, several people are involved in the decision-making process that eventually leads to personality test use in workplace training. For example, HR practitioners, line managers and departmental directors function as internal stakeholders, while coaches, psychologists and HR consultants can be considered external stakeholders (Freeman, Harrison, Wicks, Parmar, & De Colle, 2010). Test publishers and professional associations also influence the decision-making process from the outside, especially when the test publisher sends an accredited consultant to administer the personality test during a management development workshop.

According to Furnham (2004) the internal HR specialist rather than the organizational psychologist decides about which test to use. Internal stakeholders hold the training budget and they are likely to be the final decision maker for any people development initiative. It is acknowledged, however, that external consultants or coaches, some of whom are test publishers, influence the decision-making process through their personal validation of a specific test. Often regarded as “experts in the field”, external stakeholders’ advice impacts HR practitioners’ beliefs about and attitude towards tests (Furnham, 2008). The extent to which external stakeholders execute power over internal stakeholders or vice versa has not yet been explored sufficiently.

Research question 1: How do external stakeholders, e.g., publishers, associations and psychologists engage with internal stakeholders, e.g., HR practitioners, when choosing a personality test for workplace training?

Psychometric and practical considerations

When selecting “the right” personality test, stakeholders base their decision-making on different parameters. In a study among 255 UK-based HR practitioners, Furnham & Jackson (2011) highlight four parameters that influence their attitudes and beliefs about work-related psychological tests: (1) test complexity, (2) practical application, (3) bias and (4) usefulness. The study concludes that younger practitioners and those with less years in higher education tend to have a more limited understanding of psychological testing in comparison with their more educated and more experienced colleagues (Furnham & Jackson, 2011).

Other authors observe that product pricing, packaging and the publisher’s marketing materials also influence practitioners when choosing a personality test (Goodstein & Lanyon, 1999). Academic researchers, in comparison, often reference psychometric parameters such as test validity and reliability when choosing a personality instrument for research purposes (Tett & Christiansen, 2007). The question remains how HR practitioners, who are rarely experts in the field of psychometric testing, navigate between practical and psychometric considerations in their decision-making.

Research question 2: How are practical and psychometric considerations weighted during the decision-making process?

Positioning and purpose

In recruiting, personality tests are used for better personnel selection decisions (Nikolaou & Oostrom, 2015). This is not the case in HR development where personality tests' intended use is for development rather than assessment. Here, personality tests are often marketed as developmental tools that build on "easy-to-understand" models (Passmore, 2008).

In a study on cultural change in a large Australian corporation, Badham et al. (2003) find that tests are used as self-analytical tools in organizational development. The authors describe personality test use as a "rational scientific model" (Badham et al., 2003, p. 721) and find that the test performs well in situations where a discussion of feelings could have been difficult otherwise. Similarly, Ludeman's (1995) study of Motorola's HR teams across the US explores how personality tests help employees to understand what drives their behavior, and how their personality affects the way they do their job. By sharing test results in teams, an awareness of difference can help to adjust one's own behavior and to improve team effectiveness (Clinebell & Stecher, 2003).

In practice, consultants, psychologists and managers might all have different kinds of reasons for using personality testing in the workplace. It can be assumed, for example, that their own experience with testing – whether positive or negative – will be reflected in the practice of HR specialists.

Research question 3: How are personality tests positioned for the purpose of workplace training?

Methodology

Design

This study uses an exploratory research design, given the limited number of empirical studies available on the use of personality testing in workplace training. A multiple-case study design was chosen to elicit why and how this specific phenomenon occurs in different settings (Ellinger, Watkins, & Marsick, 2005; Yin,

2004). This approach seems suitable to investigate the practice of personality testing as a “contemporary phenomenon within its real-life context, especially when the boundaries between phenomenon and context are not clearly evident” (Yin, 2004, p. 13). This is relevant because of the current HRD practice that is not always driven by scientific knowledge and evidence-based management (Briner & Rousseau, 2011). Hence, an exploratory qualitative study into personality testing in its workplace training context yields an opportunity for inductive theory building.

Selection of cases

Research in personality testing tends to come from Anglo-Saxon countries, most notably the USA and the UK, but also Australia. A dominance of testing products from English-speaking countries can be observed; local language versions of these tests have been adopted throughout Europe to varying degrees (Furnham, 2004). By choosing Germany³, the UK and the Netherlands as the units of analysis, the intent is to discover personality test use in diverse Western European countries: Germany forms the largest non-English speaking market within Europe. Germany is also considered one of the “Rhineland model” markets with their more government-coordinated social market economy, sometimes also referred to as “the other capitalism” (Albert, 1993, p. 99). In contrast, the UK is heavily influenced by HR developments of other liberal Anglo-Saxon countries, like the USA and Australia. The Netherlands, quite literally, falls somewhere in between Germany and the UK in terms of its market structure and language capabilities. These three cases were selected to provide variety across a spectrum of markets in Western Europe.

Data collection

A flexible design was employed whereby data collection started in the UK and in Germany, and was later completed by data from the Netherlands. After the first version of this paper had been reviewed for publication, the authors went back to the field and collected additional data from HR practitioners in each of the country cases. Although the main data sources were semi-structured interviews, opportunities were sought to add field notes from direct observation and

³ Besides collecting data in Germany (D), also German-speaking people in Austria (A) and in Switzerland (CH) were interviewed. For simplicity, these DACH sources will be referred to as Germany in this paper.

group discussion, e-mail exchanges, websites, test reports, product brochures and other related documents to the data collection (see Figure 2.1).

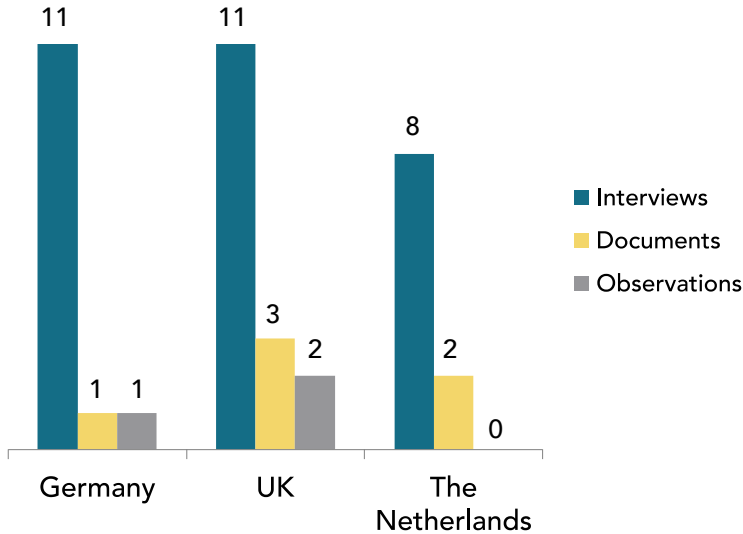


Figure 2.1 Data collection overview

Data collection ran from 2012-2016, and 39 data points were collected across stakeholders from different organizational settings, including test publishers, associations, corporate businesses, HR consultancies and educational institutions. Stakeholders were selected using convenience and snowball sampling. All interviews with Dutch and British stakeholders were held in English while interviews with German practitioners were held in German; the authors translated those transcript quotes used in this paper. At the end of every interview, stakeholders were asked for further stakeholder introductions. The authors were looking for people with different academic backgrounds, years of experience on the job and positions within their organization in order to cover a wide spectrum of stakeholders in this exploratory study. Publicly available contact information was used to approach national associations and publishers where no prior connection existed. A detailed overview of all collected materials can be found in Appendix 2.

Data analysis

Data was analyzed using an exploratory case study approach with the aim to gain deeper insights into the dynamics of the personality testing industry. The researchers went on to read and code the collected data several times. During the first round of coding, all accounts were organized chronologically per country with the aim to distil a descriptive narrative that would be closely aligned to the research questions. Next, and during the second read, this descriptive narrative was pulled apart in different ways to highlight salient topics per case. The aim here was to use Yin's (2004) methodology of pattern matching to get underneath the story and to depict dynamics and tensions within the industry. Finally, a last round of reading and coding was performed to integrate insights and knowledge from different kinds of sources through triangulation; this was done by comparing and matching statements in a sense-making way. The aim here was to interpret and explain the findings from within cases, and to come up with emergent themes across cases along Eisenhardt's (1989) approach to theory-building based on case studies. These themes can be seen as pieces of the jigsaw puzzle and a first step towards meaning-making and theory-building of personality testing in workplace learning.

Findings – country context

The findings sections starts with an overview of each case study context, in which national associations, standards and reports relevant to personality testing in that country will be highlighted.

Personality testing in Germany

In Germany, the Test Review System of the Board of Assessment and Testing (TBS-TK), supported by the Professional Association of German Psychologists (BDP), helps psychologists and non-psychologists to choose the right test. The TBS-TK publishes test review reports that are based on a norm called DIN 33430 (DIN, 2002). The norm, related to the standards laid out by the European Federation of Psychologists' Associations (EFPA), defines and details proficiency in assessment procedures. Other national associations and testing centers in Western Europe have adopted similar test standards for test reporting (see Table 2.1).

Table 2.1 Associations and testing centers

Category	Organization	Acronym	Country	Description
Associations	Professional Association of German Psychologists	BDP	D	www.bdp-verband.org
	British Psychological Society	BPS	UK	www.bps.org.uk
	Chartered Institute of Personnel and Development	CIPD	UK	www.cipd.co.uk
	Dutch Association of Psychologists	NIP	NL	www.psynip.nl
	Dutch Association for Personnel Management and Organization Development	NVP	NL	nvp-plaza.nl
	European Federation of Psychologists' Associations	EFPA	EU	www.efpa.eu
Testing Centers	Test Review System of the Board of Assessment and Testing	TBS-TK	D	www.zpid.de/index.php?wahl=Testkuratorium - all test reports are publicly available
	Psychological Testing Centre	PTC	UK	ptc.bps.org.uk - in order to access test reports you have to be a member of BPS
	Dutch Committee on Testing Affairs	COTAN	NL	http://www.psynip.nl/tests_cotan.html - in order to access test reports you have to be a member of NIP

While the DIN 33430 norm for German speaking countries does not detail specifically the use of personality testing in a learning context, it applies to everyone, irrespective of the profession of the assessor (Hagemeister, Kersting, & Stemmler, 2012). This is an important aspect as 80-90% of psychological tests are sold to non-psychologists who come from “a wide range of different backgrounds in terms of skill and experience” (Ackerschott, 2000, as quoted in Bartram, 2001). However, the adoption level of the norm is not clear as “people

have heard of DIN 33430 and TBS-TK but they do not necessarily know its content” (Hagemeister et al., 2010, p. 438).

At the time of writing, 30 TBS-TK test reviews were publicly available online, six of which were reviews of personality tests⁴. Of these six reviews, four resulted in a “pass” and two in a “less suitable” test result. Whether those test review results directly impacted the marketability of each personality test could not be established.

In July 2014, the German consumer organization Stiftung Warentest – equivalent to *Which?* in the UK or *De Consumentenbond* in the Netherlands – published a personality test report. Titled “What am I? Personality Tests Online” the report reviews ten instruments (all publicly available for individual test takers) along several criteria that were again based on the German DIN 33430. The verdict: None of the ten tests got a “sehr gut” (very good - using the German school’s marking system), only two of the ten tests are considered “good”, half of the tests scored “satisfactory” or “adequate”, and three tests even “failed” the standard (test.de, 2014).

In Austria, the ÖNORM D 4000 that gives guidelines for standards and processes in personnel selection and personnel development has been implemented (Österreichisches Normungsinstitut, 2005). Educational professionals in the German speaking part of Switzerland also discuss standards in test use along the DIN 33430 norm (Felder, 2005).

Personality testing in the UK

In the UK, the Psychological Testing Centre (PTC) is the body within the British Psychological Society (BPS) that evaluates MBTI, Belbin, OPQ, TMP and similar tests against psychometric testing standards (see Table 2.2).

⁴ At the time of writing, these six personality test reviews by TBS-TK were: BIP-6F; FPI-R; Golden Profiler of Personality GPOP; NEO PI-R; OPQ32; Persolog Persönlichkeitsprofil.

Table 2.2 Popular personality tests in Germany, the UK and the Netherlands

Category	Test name	Publisher/ Distributor	Mainly used in	Test reviews by ⁵
Trait	NEO PI - Neuroticism, Extraversion, Openness Personality Inventory	Hogrefe	D, UK, NL	TBS-TK (D) PTC (UK)
	OPQ32 - Occupational Personality Questionnaire	SHL Talent Measurement	UK, NL	TBS-TK (D) PTC (UK)
	HPI - Hogan Personality Inventory	Metaberatung (D); Psychological Consultancy (UK)	D, UK	PTC (UK)
	16PF® - Sixteen Personality Factor Questionnaire	OPP	UK	-
	BIP™ - Business-focused Inventory of Personality	Hogrefe	D	TBS-TK (D)
Type	Management Drives	Management Drives	NL	-
	MBTI® - Myers-Briggs Type Indicator (Step One)	OPP	D, UK, NL	PTC (UK)
	TMPQ - Team Management Profile Questionnaire	TMSdi	UK	PTC (UK)
	IDPE - Insights Discovery Preference Evaluator	Insights Group Deutschland	D, UK, NL	PTC (UK)
	Belbin Team Roles	Belbin Associates	UK	-

The PTC categorizes different personality tests, which are available to members of the BPS. At the time of writing, 24 test reviews were found in the category “personality – trait” and 6 test reviews in “personality – type”. Some tests fell into two categories, such as the “Jung Type Indicator” that was classified as both *type* and *trait* tests.

Furthermore, the Chartered Institute of Personnel and Development (CIPD) plays a role in connecting test publishers and HR practitioners. The CIPD issues

⁵ Only the German TBS-TK and the British PTC/BPS test reviews could be accessed as part of this research project. Access to the Dutch COTAN database was more difficult to negotiate as it is for registered members only.

an annual report on latest learning and development (L&D) trends. Published since 1998, the report included a section on “Individual and Team Learning Analysis and Diagnostics” (McGurk & Belliveau, 2012, p. 21) for the first time in 2012⁶. A target population of 21,222 people was approached with a survey instrument consisting of 51 questions. L&D experts from the public and private sectors were asked about their use of diagnostic tools, and the report found that “many organizations use one or more methods of learning analysis/diagnostics” in their practice. For example, the MBTI was used by more than “two-fifths of those who responded to this question” and Belbin Team Roles, a team assessment questionnaire, was used at least “occasionally” by nearly half of the organizations (McGurk & Belliveau, 2012, p. 21).

Personality tests are also discussed and debated in more public contexts. For example, in a joint effort between BBC Lab UK and the University of Cambridge, people were asked to complete an online survey between 2009-2011 (Rentfrow, Jokela, & Lamb, 2015). The so-called “Big Personality Test” is based on the Big Five personality model⁷, and more than half a million Brits completed this trait test. Although the test results are not individualized, initiatives like this “Big Personality Test” expose the public to the topic of traits and testing in a scientifically controlled and enlightened way.

Personality testing in the Netherlands

In the Netherlands, the Dutch Committee on Testing Affairs (COTAN), a subsidiary of the Dutch Association of Psychologists (NIP), is a testing center that looks after quality standards in test use. The COTAN also assesses tests and provides reports on their suitability as a guide for psychologists and test users (Evers, 1996). These test reports are hosted in the members-only section of the COTAN website, accessible by registered psychologists with NIP membership. The Dutch Association for Personnel Management and Organization Development (NVP) is another organization available to the industry with a focus on HR topics, but with no specific expertise in personality testing. Other publicly

⁶ The special section on “individual and team learning analysis and diagnostics” appeared only in the 2012 report, but not in any of the following reports in 2013, 2014 or 2015. The 2016 report was not yet available at the time of publication.

⁷ The BBC Lab UK questionnaire asked 44 personality-related questions, whereas the full length NEO PI consists of 240 questions and the shortened NEO PI-R of 63 questions.

available sources that discuss personality test use include Vermeren’s (2009) book *De HR-ballon* (“The HR Balloon”, published in Dutch). In this book, ten common HR theories, models and tools are depicted that – according to the author – are “nothing more than popular pseudo-science” (Vermeren, 2009, p. 5) as they lack the empirical foundation. Two chapters of the book focus on personality testing, with an emphasis on the pitfalls of type tests such as the MBTI. The book is practitioner-focused and easy to understand by people working in management or HRD.

Findings – emergent themes

Next, the remaining findings will be laid out along four emergent themes: tensions, decision-making considerations, purposes, and concerns. Each findings section presents quotes that support the evidence. References are made to the various stakeholders interviewed and places visited in this case study research.

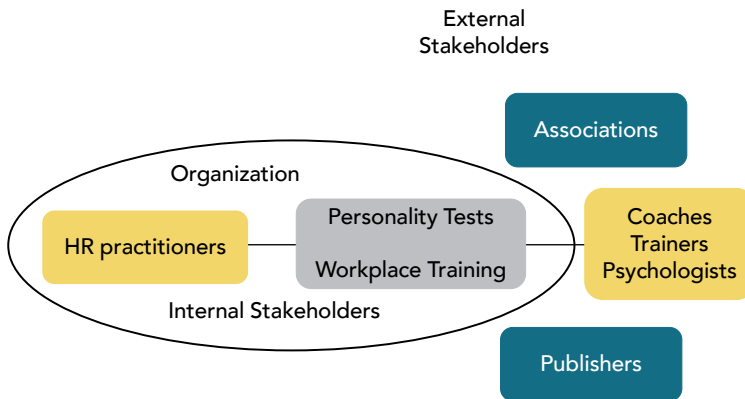


Figure 2.2 Stakeholders in personality testing industry

Tensions: Psychologists, non-psychologists and associations

Several stakeholders are involved in the personality testing industry. Within the organization, there are HRD practitioners whose objective is to run successful development initiatives, for example, in the form of a workplace training using a personality test (focus of this study). On the outside, there are independent coaches, trainers and psychologists who support the organization with running

workshops and other development initiatives. Other external stakeholders include test publishers and professional associations (see Figure 2.2).

An overview of professional organizations in each country case was presented earlier in this paper (see Table 2.1).

The CIPD HRD Exhibition in London creates ample opportunity to pick up test brochures and to talk to test publishers in an attempt to understand the position of publishers in the personality testing industry (Doc011). In the UK, there are four large publishers that dominate the market (see Table 2.3). Using semi-structured interviews, each publisher’s market position was explored, with a special emphasis on anticipated challenges.

2

Table 2.3 Overview of test publishers with challenges

Publisher	Challenges	Source
OPP Ltd (Oxford)	<ul style="list-style-type: none"> • Economic climate: coming out of recession • UK more saturated than other European markets • Cultural difference, e.g., US population more used to test taking 	Doc010
TMS Development International (York)	<ul style="list-style-type: none"> • Competition from other test publishers, e.g., OPP Ltd • Expectation from internal stakeholders that tests can be fully explored in 2h or 4h team development events 	Doc001
SHL Talent Measurement (London)	<ul style="list-style-type: none"> • Competition from other test publishers, e.g. Saville Consulting • SHL tool OPQ not as applicable for senior management 	Doc008
Hogrefe UK (Oxford)	<ul style="list-style-type: none"> • Little knowledge of Hogrefe tools (NEO PI-R and BIP) among HR practitioners 	Doc015

Economic climate, competition and varying degrees of test popularity were named as challenging. Notably, testing culture was mentioned as a differentiator between the US and European countries: since Europeans are less used to standardized school testing in general, this would also negatively impact adoption levels when it comes to tests in workplace training initiatives (Doc010).

In the Netherlands, skepticism exists towards test publishers and consultancies that “ride the waves and capitalize on the enthusiasm of the public” (Doc019b). From the perspective of the external psychologist, publishers pose a threat to

professional practice. “The worse the quality of the instruments, the better their marketing is,” (Doc020, lines 214-215), comments a registered psychologist from the Netherlands who has more than 35 years of experience in the use of personality testing.

Because of quality concerns, the Dutch association NIP started to intervene against the practice of non-psychologists administering tests: all registered psychologists who were training non-psychologists in test use would lose their membership to the association. “The NIP decided to kick out these psychologists” (Doc018, lines 64-65). Other psychologists also commented on their relationship with the NIP that made them ultimately leave the association, for example:

I was a member of NIP, but I got a little bit of disagreement with them about something. I thought that they were giving too little real support to psychologists. I didn't see the added value of them as an organization, so I went out (Doc022, lines 165-168).

In a similar way, tension exists between the British association BPS and HR practitioners: “...the BPS is effectively the sort of governing body, really, of test validity and reliability, that sort of the organization I'd probably go to, even though I do find them a very arrogant sort of organization” (Doc032, lines 315-318). With a background and Masters degree in organizational development, this UK practitioner does not qualify to become a member of the national psychological association. Generally, a psychology degree is mandatory for national membership in one of the associations, which makes it hard for non-psychologists to get access to test reports and latest research evidence.

The relationship between practitioners and the national psychological associations is a complex one: On the one hand, the BPS and the NIP are the “go-to” places for psychometrics, but on the other hand they are not very supportive of HRD practitioners and their work. When asked about the test validation reports issued by national testing centers, the responses were along this line: “There are hardly any customers who know that something like the COTAN exists” (Doc020, lines 182-183). Consequently, test reports are less frequently consulted by non-psychologists and rarely demanded by HR practitioners (Doc020).

An experienced psychologist from the Netherlands explained this by using the “going to the dentist” analogy when it comes to the test selection criteria: “I am interested in his quality as a dentist but not so much in the quality of the instruments he is using” (Doc020, lines 192-193). Attempts to request an opening-up of NIP membership to “laymen” have been fruitless so far (Doc031, lines 294-302).

In Germany, psychological associations and membership are mentioned less, but that does not mean that people do not feel strongly about who should administer personality tests at work. “This is the field of psychologists, and, yes, other people can do it, but not in my organization” (Doc027, lines 229-230). This experienced practitioner with a PhD in psychology uses the analogy of going to the doctor (not the dentist this time!): “If you are sick, you go to a doctor for medicine, medicine man, and not to a carpenter ... There is a field of study which is responsible or educated in this” (Doc027, lines 237-239). During the conversation, this highly qualified psychologist considered whether someone trained in a specific personality test should be able to administer just that one test – analogous to a nurse drawing blood and administering a specific treatment. After some further reflections, the psychologist nevertheless remains hesitant about non-psychologists’ practice.

In comparison, a non-psychologist practitioner from Germany does not see a problem in administering tests without a psychology degree. Even stronger, the HR director of a multinational consumer goods company regards test use as part of her role as an HR business partner: “It is our aim that the entire [HR] team gets licensed, so that they can administer the tests themselves ... I think that this must be part of our skill set and our self-concept” (Doc026, lines 218-219; 233-234). The tendency to administer personality tests internally seems to be supported by a younger generation of managers and leaders who are described to have a heightened awareness of certain topics: “The general knowledge around those [leadership] topics has increased ... generally, the common manager nowadays knows a lot ... including leading through mentoring ... and leading through coaching” (Doc021, lines 407-410; 414-415).

Another German HR practitioner and consultant with years of experience in various industries added that since the translation of Daniel Kahneman’s “Thinking, Fast and Slow” and the book’s vast distribution through airports,

German business literature readers are much more familiar with business concepts and topics imported from the USA. “In a similar way, with some time lag, personality tests have simply also arrived” (Doc024, lines 370-371).

In sum, findings show that tension exists between national associations and practitioners, especially where test users without a psychology degree do not have a “go-to” association when it comes to test selection and best practice in test use. Since non-psychologists are less “uniform” in their test use and application of, for example, type tests, some psychologists feel bitter about their practice. HRD practitioners, on the other hand, regard test use as part of their skill set and self-concept.

Considerations: Ease-of-use, trait-tests and historical choices

Personality tests are chosen for workplace training based on different considerations. When asked, “How do you choose a personality test for a given workplace training?” some HR practitioners mention psychometric standards, such as validity and reliability of the tool, as part of their decision-making process (Doc004; Doc12a; Doc027; Doc032).

However, practitioners also talk great length about other criteria that are important when they make the decision about which instrument to choose (Doc021; Doc024; Doc026; Doc030; Doc033; Doc034). For example, an experienced HR leader from the UK explains, “I’m not a fan, for example, of Myers-Briggs ... [the] reason being [that] while it’s a very valid tool in my mind, it’s a little hard for people to get their heads around and apply.” (Doc034, lines 69-70; 72-73). A consultant based in Germany adds, “The DISC is so simple and you can easily remember the colors; I don’t mean it in an unkind way but my colleagues are not willing to remember more than four colors.” (Doc024, lines 293-295). So here, the perceived ease-of-use of the tool from the test users’ perspective plays a role in their decision-making.

For some practitioners, the choice of the personality test is a historical choice: Their employer organizations started with one tool some time ago, and then kept it. In that case, practitioners joining the organization later in time do not know exactly why that tool was used, as they had simply “inherited” it: “...the psychometric of choice in this organization is Insights, and I’m not an accredited

practitioner, but I do have two colleagues who are.” (Doc032, lines 80-82). When asked, the UK-based practitioner does not know the exact criteria that led to choosing Insights as the organization’s tool of preference; the practitioner can only point towards the fact that it has been used with her employer for a longer period of time – no further questions asked.

For other practitioners, there is a clear preference for one set of tools over another: “To be honest, I like the Big Five. I’m not so much a type questionnaire person. ... Again, I feel that when you work with typologists ... you put someone in a certain color range, and that’s it” (Doc028, lines 83-84; 93-94). Especially when speaking to psychologists or more experienced HR practitioners, a clear line between different personality test approaches is drawn:

*It’s what I call the *kleur terreur*, the “terror of color” because there are the Management Drives, the DISC and others like that. They are all colors, and you can be a red person or a yellow or orange. That’s what they do... that’s also what I have against the MBTI because it’s labeling people (Doc031, lines 131-133; 134-135).*

Here, the practitioner eyeballs critically type and prefers trait tests; these are tests that refrain from using colors or categories to describe the test taker. As one critical HRD practitioner from Germany explains: “I think there’s a lot of over-the-counter type testing, that actually do more damage than good” (Doc023, lines 375-376).

The topic of pigeonholing comes up in most conversations, and practitioners seem to employ different strategies to overcome it in their own practice. One strategy is to avoid type tests and to administer trait tests only. “If we take OPQ, for example, there is no label there” (Doc025, line 276). Another way of dealing with the potential risk of pigeonholing is to introduce the type test in a very soft way, as a “conversation starter” (Doc033, line 232) rather than a tool that gives you a definite answer about yourself and others. Yet another strategy is to administer multiple tests at the same time or to combine the personality test with other tools: “we conducted a 360 and a Hogan in combination” (Doc027, line 31).

Finally, the positioning of publishers and their way of interacting with HR practitioners also influences the decision-making process. Personality tests tend to be sold in “bundles”, which means that they come with test reports of various lengths and forms, presentation slides, training aids, etc.: “...they’re very keen to make sure that you have as much support as possible in terms of using the tool and understanding it” (Doc032, lines 154-155). This specific test publisher is characterized as presenting itself in a “consultative” way, meaning that it provides additional support services in order to help the test administrator to use the product in an effective way – at no additional charge. However, not all publishers are described as equally valuable: “They’re very professional, but they’re quite hands off. I always get the impression that if I wanted a bit more insight or support then that would come at a cost” (Doc032, lines 171-173).

In a more extreme case, test publishers encourage HR practitioners to turn into “sales people” – according to a report of a UK consultant, who started his career as an internal trainer and later went independent:

I don't feel it's ethically right that I can get the profiles for 50 Pounds and I charge the client 90 to 150. I just think that's a bit naughty but they do it because that's part of how they fund it plus you pay a yearly license fee (Doc030, lines 87-90).

In sum, HR practitioners consider different parameters when choosing from the many personality tests that are available on the market, including cost, perceived ease-of-use, trait or type approach of the tool, as well as the relationship quality with test publishers. Psychologists in this sample strongly voice their opposition to type test use. The availability of official test review reports, as issued by the national testing centers, plays a marginal role in the decision-making process. In some instances, however, the test-choice is a historical one; in this case, the HR practitioner has not been involved in weighing the different practical and psychometric considerations – and non-psychologists seem to be at ease with accepting such historical choices.

Table 2.4 Purposes for using personality tests

	Individual	Team
<i>Exploratory</i>	<p>Purpose 1: To encourage individual reflection</p> <p><i>"[A] personality questionnaire is something I use to start a reflection process with people, and to open a dialogue." (Doc028, lines 66-67)</i></p>	<p>Purpose 2: To explore differences within teams</p> <p><i>"For people to see, to notice, to take serious and to accept differences and sameness." (Doc021, lines 185-186)</i></p>
<i>Developmental</i>	<p>Purpose 3: To initiate personal development</p> <p><i>"To collect ideas on how to change behavior...awareness is good, change is better." (Doc024, lines 120; 117)</i></p>	<p>Purpose 4: To fix dysfunctional teams</p> <p><i>"It may be that there has been a breakdown in relations in the team and it's used as a "soft way in" to try to explain why the breakdown might have happened and start to remedy that way." (Doc033, lines 23-25)</i></p>

Purposes: Exploratory or developmental

HRD practitioners describe different purposes of personality test use in developmental contexts, none of which is actually to “measure” personality. Practitioners are eloquent in stating that they use those assessments with purpose. The use of personality testing in the workplace differs both in terms of scope (individual versus team) as well as intent (exploratory versus developmental). A simple four-box table describes the resulting four categories of test use purpose in workplace training (see Table 2.4).

Some interventions focus on the individual employee, whereby a questionnaire is administered as part of a leadership development program or an individual coaching setting. Other interventions include the whole team when working on team development. The intent of those individual or team interventions differs, too: from exploratory interventions on self- and team-reflection to more developmental agendas wherein fixing a dysfunctional team or helping a manager to adjust behavior for better performance results could be intended learning outcomes. Notably, team profiles such as Belbin Team Roles and the Team Management Profile Questionnaire are more often mentioned by HR practitioners

from the UK country context (see also Furnham, 2004 who gives reasons why team profiles are stronger in the UK).

To conclude, practitioners have different purposes in mind when administering personality tests in workplace training contexts: the focus can be individual or team-based, and the intent can vary between exploration and development.

Concerns: Pigeonholing, privacy and re-labeling

When asked about their personality testing practice, practitioners voice different concerns that could – if left unmanaged – inhibit the intervention from being successful. Through their practice and along the guidelines of test publishers, practitioners have developed mechanisms to address these common concerns, including the concern of pigeonholing as detailed previously. One HR practitioner from the UK who now lives abroad explains how the concern is two-fold: first of all, practitioners could use the assessment in a categorizing way, and secondly, participants in the intervention might do so, too: “You really have to watch for people ‘pigeonholing’ themselves as in ‘I’m an ISTJ’. I encourage people to look beyond this and stress that these are preferences” (Doc033, lines 58-59).

Another concern centers on privacy. Here questions as to who should receive the test feedback and how the results should be stored are answered. Interestingly, not many practitioners talked about privacy policies when it comes to handling tests in the workplace. Is this a topic taken “for granted” and hence little discussed? The one German psychologist who did mention the company’s internal testing policy had a background in recruiting, where explicit assessment policies are more common.

The word “test”, especially when used in the German context, poses another concern: “Because Germans react negatively to the word ‘test’” (Doc024, lines 76-78). The consultant explains that Germans do not like the word “Kontrolle” (control) either but do not mind the word “monitoring” instead. He further explains how practitioners tend to “re-label” tests in order to “disenchant” them from their negative connotation. For example, an education consultant at a school of adult education in Vienna used the term “Potenzialanalyse” (analysis of potential) when personality tests were embedded in coaching sessions (Doc009). Practitioners explain that skepticism towards the use of personality

testing can be lowered if a more neutral, and “less privacy-invasive” (Doc003) term is used. Similarly, the terms “Standortanalyse” (situational assessment) and “Standortbestimmung” (personal stocktaking) appear more often in the adult education and HRD contexts in Germany (Doc007 and Doc016). So using “the appropriate” terms and hence “re-labeling” tests in a given context is a technique that practitioners adopt to make personality tests more appealing to broad audiences.

Moving down the list of concerns, employees prefer to look at the “bright” side of their personality, which decreases the use of tools that also focus on traits like “neuroticism” (NEO PI-R) or “de-railers” (Hogan). “People don’t like to see negative things about themselves often ... there may be 95% positive things in there, and 5% negative, and those 5% are what they will focus on” (Doc023; lines 159-162). This statement by a German practitioner with a background in organizational behavior is interesting as it shows another feature of personality test marketability: the more positively-phrased assessments get more sales uplift as they confirm what people like to talk about – that is: themselves from a positive standpoint.

The same is true for marketing materials that do not differ greatly among tests, and that are mostly phrased in a very positive way. When speaking to publishers, the NEO PI-R and the MBTI were often referred to as the two tests on opposite ends of the spectrum (Doc001; Doc008; Doc010; Doc015). However, reviewing the marketing materials from both publishers shows little significant difference between the two assessments (see Table 2.5).

Table 2.5 Comparison of NEO PI-R and MBTI marketing leaflets

	NEO PI-R	MBTI
UK Publisher	Hogrefe UK	OPP Ltd
Category	Trait	Type
Purpose	“Deepening understanding of people at work”	“To gain a deeper understanding of themselves and how they interact with others”
Application (Top 3)	Selection Development Vocational guidance	Team development Leadership development Communication
Key features	<ul style="list-style-type: none"> • 240 questions • Takes around 30 minutes • Paper or online, numerous languages • 5 broad personality areas • Automatically generated report 	<ul style="list-style-type: none"> • 16 easy-to-understand types • Available online in many languages • Outstanding credentials, backed by more than 6,000 independent research papers • Comprehensive application reports
Benefits	<ul style="list-style-type: none"> • Looks beyond past achievements • Gets to the root how this person works: working style, team behavior, how conflict is resolved • One of the gold standards of personality assessment for the business market 	<ul style="list-style-type: none"> • Lays a positive foundation of type knowledge that can be applied • Gets to the root of difficult development challenges, provides understanding of the whole person • Delivers quality-assured results that address a wider range of business issues
Norm groups	UK working population, UK job applicants, financial services	NA
Testimonial	“NEO provides a deeper and fuller picture of an individual’s personality than some other tools...” Tim Evans, Psysoft	“After qualifying, it was very easy to apply the training by using the online OPP assessment system to generate detailed MBTI report.” Tim Evans, Psysoft

Whilst OPP promotes its tool as “16 easy-to-understand types”, the number of questions that would have to be completed for the NEO PI-R (240 in total) seems rather time-consuming. Surprisingly, the same person (Tim Evans from Psysoft) gives testimonial for both competing products. For the rest, similar language, structure and benefits are stated, which makes selecting a test based on marketing materials quite difficult.

According to a senior Dutch psychologist who administers tests inside client organizations, the need for validation is high for test takers: “It’s always a surprise to me that people recognize themselves” (Doc019a, lines 99-100). According to the observation of this practitioner, it does not seem to matter how statistically robust the test is, as positive wording of relatively general statements seems to speak to most people according to the observations of this practitioner. Linked to self-validation is the concern that test feedback could be perceived as an end rather than a starting point of a conversation. “Some people love psychometrics; it almost becomes a gospel to them. Lots of people become almost evangelical about what they learnt” (Doc034, lines 244-245). This idea that someone fully adopts the feedback and becomes almost religious about it can lead to alienation within teams and among colleagues: “If you get a leader running around saying, “I’m ENTJ,” you look at him and go, ‘Stop talking crap. Talk to me like you’re a real person’” (Doc034, lines 273-275). Here, the positive attraction of the test feedback for one employee can have a negative impact on a colleague who is more reserved about the practice of personality testing in the workplace.

In sum, many practitioners who use tests on a regular basis also reflect on specific concerns, such as pigeonholing, privacy and re-labeling the term “testing”. All practitioners develop strategies on how to mitigate the risk of the tests being perceived in an unintended way.

Conclusions and discussion

This case-study research explored the use of personality testing in Western Europe to clarify why tests are used in workplace training and how they are implemented. The exploration was guided by three research questions that looked at the involvement of internal and external stakeholders who weigh

psychometric and practical factors when explaining personality test use for workplace training.

RQ1: Stakeholders

The first research question asked about external stakeholders and their interaction with HR practitioners when it comes to choosing a personality test for workplace training. This exploration shows that decisions are initiated and made by HR practitioners themselves if specific test use has not been established already. Some practitioners closely work with consultancies and external partners who often have preferred personality tests in their portfolio; others maintain direct links to test publishers with whom they have worked in the past. Some practitioners prefer to do everything in-house and only purchase the specific test via a distributor.

Psychological associations and test publishers have a much smaller decision-making influence on HRD practitioners than previously established (Furnham, 2008). A void exists between associations trying to regulate the testing industry and psychologists and non-psychologists who apply personality tests in their professional practice. Some practitioners feel disillusioned about the non-support of their associations – as could be seen in example of two Dutch psychologists who decided to drop their NIP membership. For other practitioners, membership in psychological associations is denied, which makes it hard for this growing group of HR practitioners without a psychology degree to retrieve test reports and base their decision-making on standard test reviews (Hagemeister, Lang, & Kersting, 2010). As non-psychologists have less professional membership options and no clear “go-to” association, test publishers gain influence in their decision-making process, which further expands the void between psychologists and non-psychologist practitioners (Evers et al., 2012).

The findings also confirm that the lines between external and internal stakeholders are very fluid: practitioners are not stuck in one career. For example, one psychologist interviewed used to work for a test publisher and later became a consultant. Another interviewee was a consultant before getting hired into an HR management position. Similar to career changes, the decision whether to insource or to outsource goes in cycles. Many organizations in this sample

had seen both: the use of external consultants as well as the strategy to train up internal personnel in test usage.

RQ2: Practical considerations

The second research question looked at the weighing of practical and psychometric considerations during the decision-making process. HRD practitioners consider cost and test complexity as two important factors. Perceived ease-of-use trumps psychometric factors. Some practitioners are stuck with a tool for historical reasons. They continue using tools even if they are perceived as “un-fit” to achieve the intended purpose (Kuipers, Higgs, Tolkacheva, & de Witte, 2009).

HRD practitioners are generally aware of the type test criticism, whether they are psychologists or not. However, non-psychologists employ different strategies to continue using type tests where they see fit. Psychologists tend not to do that but reject type tests on the grounds of their reported psychometric limitations (Boyle, 1995; Pittenger, 2005). The psychologists’ preference is to use trait tests, such as the NEO PI-R, Big Five, OPQ or Hogan, in their practice. It was surprising how often Hogan tests were mentioned as a good alternative to other more popular tests, despite their distinct and more negative feedback language, e.g., using the term *derailers* to describe “the dark side of personality” (Nelson & Hogan, 2009; Spain et al., 2014).

RQ3: Purpose and concerns

The third research question centered on the purpose and positioning of personality tests in workplace training. The findings show that HRD practitioners vary their purpose depending on scope and intent; some focus on teams, other practitioners focus on individuals; some employ a light touch of exploration, other personality tests are used with the clear intent of change and development. An apparent paradox arises when the purpose to change is compared with the fact that personality is said to be relatively stable over the course of your life (Srivastava et al., 2003). How can something be stable and changeable at the same time? The word *behavior* seems key in this context to explain what a personality test enables in a developmental setting: by receiving feedback on personality traits, preferences and likely behavior when acting within preference, the individual or team is reminded that behavior is changeable, and through changes in behavior,

personal development and growth can be achieved. This behavior change is possible even if it may not suit the person's preferences where employees with a higher level of self-awareness can respond more proactively to their external environment (McCarthy & Garavan, 1999).

Furthermore, HR practitioners are aware of concerns and criticism that surround test administration and test use (Pittenger, 2005), and they have built mechanisms into their practice to prevent them from being perceived as negative. For example, practitioners re-label tests as tools for *self-reflection* and instruments for *personal stocktaking*. Testing and assessment language disappear from marketing materials and jargon of potential and growth gets introduced, in line with the developmental purposes that can also be observed with other feedback tools (McCarthy & Garavan, 2001).

Research implications

Evidence-based management is something done by practitioners, and not by researchers. This paper shows how critically-reviewed research evidence can be complemented with practitioner experience and judgment, evidence from the local context and the views of those stakeholders who are affected by the practice. By adding lived experiences and held beliefs to the picture, nuances of test use become visible, and individual and organizational decision-making trade-offs transpire. Not all practitioners fuel their decisions with research evidence; other influences such as the tool's historical use within the organization and the influence of external stakeholders were highlighted.

In their cross-cultural study on the research-practice gap in human resource management (HRM), Tenhiälä et al (2016) found that the surveyed HR practitioners "are not evidence-based as their knowledge level of the research evidence on effective HRM practices can only be described as 'fair'" (2016, p. 193). The researchers had asked practitioners from Finland, South Korea, Spain, the United States and the Netherlands to confirm statements on HR research evidence as either *true* or *false*. Although the present study did not evaluate practitioners' knowledge of personality testing evidence, it became clear that many practitioners from our sample might have scored low grades on a true/false exam on personality testing research evidence. We need to acknowledge that restrictions on access to knowledge exist in the personality testing industry, and that

scholars, educators and psychologists can all play an important role in building support structures for the practice of evidence-based management (Briner et al., 2009; Briner & Rousseau, 2011). As Reay et al (2009) write: “the goal of evidence-based medicine is that physicians adopt practices most likely to improve patient outcomes” (2009, p. 13). Similarly, the goal in evidence-based HRD should be to adopt practices most likely to improve organizational outcomes. Knowledge transfer in the form of more systematic literature reviews needs to fill the gap as well as develop ways to bring evidence to practicing managers, consultants and organizational psychologists.

Research limitations

One limitation of this study is its weak “replication logic” implementation (Yin, 2004, p. 34) where in each country case should have followed the exact same data collection pattern. In this present study, however, stakeholders were interviewed from various organizations and at different times across the three cases and multiple years, which limits the possibilities for cross-case comparison. Every time that a contextual condition, such as a country-specific testing norm, was explored, more questions came up, some of which were investigated in one country but not in another. For future research, an improvement could be to start from a “case of practice” where a personality test is used within a specific workplace context in one or multiple locations. For example, in a case of practice with Motorola, Ludeman (1995) researched how HR professionals engaged in a team development initiative using two different personality tests. Starting from a workplace training initiative (rather than an industry), researchers could observe the dynamics of different stakeholders; their decision-making considerations and the explanation of scope, purpose, and intent in a future case of practice study.

The findings also show that some case-specific variation exists among Germany, the UK and the Netherlands, especially when it comes to the country dynamics with national associations. Despite elaborate data collection in various formats over a number of years, this study focused on only three cases from Western Europe. The extent to which industry tensions, decision-making considerations and re-labeling can also be found outside this geographic area remains unclear: Do French test publishers hold an equally powerful position when advising their business clients as their Dutch counterparts do? Does re-labeling of tests also

take place in non-European countries, and if yes, what are the concrete terms that substitute the term “testing”? For future research, a survey instrument could be developed and administered internationally to test some of these possibilities for generalization, similar to the national study conducted by Furnham (2008), who looked at HR professionals’ beliefs about and knowledge of psychometric testing in selection and development in the UK.

Lastly, the test takers’ search for validation should be looked at with more care: is it true that test takers are happy to accept any type of personality test feedback—commonly referred to as the Barnum effect (Tobacyk, Milford, Springer, & Tobacyk, 1988)—no matter how accurately it represents them as a person? It would be interesting to verify employees’ and managers’ position as test takers, who could be interviewed in a follow-up study.

Practical implications

Some efforts can already be observed as first steps towards more evidence-based HRD practice around personality testing. For example, BBC’s “Big Personality Test” (Rentfrow et al., 2015), Vermeren’s (2009) book *De HR-ballon* and the German Stiftung Warentest’s (2014) “Testing the Test” report all play a role in disseminating test quality information that could form the basis of better decision-making in HRD practice.

Since employees and managers are likely to be confronted with personality test use for development at some point in their career, HR practitioners’ increased awareness of the dynamics and challenges in this field may lead to more employee-oriented outcomes. There is a very practical implication attached to the wider use of personality testing: since non-psychologists have no independent “go-to” organization to educate themselves about test use and best practice, it is no surprise that HRD practitioners lean to peers and test publishers to get help with their own practice. Publishers have realized this gap, and they now actively offer their help and support to clients. Especially those publishers who use a consultative approach and “business partnering” are seen as highly valuable from the HRD practitioners’ perspective. Researchers, educators, and industry stakeholders should look critically at this trend where commercial publishers give guidance to practitioners instead of associations who used to safeguard testing ethics. It is the shared responsibility of the HRD community to challenge

personality tests and their use by offering reflections and an open discussion platform for more evidence-based practice going forward.

Chapter 3

On critical reflection

A review of Mezirow's theory and its operationalization⁸

⁸ This chapter has been published as: Lundgren, H., & Poell, R. F. (2016).
On Critical Reflection: A Review of Mezirow's Theory and Its Operationalization.
Human Resource Development Review, 15(1), 3-28.



Abstract

In this paper, we review empirical studies that research critical reflection based on Mezirow's definition. The concepts of content, process and premise reflection have often been cited, and operationalizing Mezirow's high-level transformative learning theory and its components has been the endeavor of adult education and human resource development (HRD) researchers. By conducting a literature review, we distill 12 research studies on critical reflection that we dissect, analyze and compare. Discovering different approaches, assessment processes and outcomes leads us to the conclusion that there is little agreement on how to operationalize reflection. We suggest four improvements: (1) integrating different critical-reflection traditions, (2) using multiple data collection pathways, (3) opting for thematic embedding and (4) attending to feelings. By implementing these improvements, we hope to stimulate closer alignment of approaches in critical-reflection research across adult education and HRD researchers.

Introduction

To reflect critically about one's own practice is often seen as the starting point for gaining new perspectives in the daily routines of working professionals. Whether you look in leadership development (Li, Gray, Lockwood, & Buhalis, 2013; Muir, 2014; Sherlock & Nathan, 2004; Nesbit, 2012), in management and entrepreneurial learning (Cope, 2003; Franz, 2010; Gray, 2007), in medical and health professional education (Keevers & Treleaven, 2011; Mann, Gordon, & MacLeod, 2009; Wald, Borkan, Taylor, Anthony, & Reis, 2012), in higher education (King, 2004; Kitchenham, 2006; Kreber, 2004; Kreber & Castleden, 2009), in correctional institution education (Behan, 2007), or in the learning of teams (Schippers, Den Hartog, & Koopman, 2007) and individuals in the workplace (Van Woerkom, 2004, 2003), the practices of *critical reflection* have been researched in many different areas of the human resource development (HRD) and adult learning spectrum.

John Dewey (1933), the founding father of reflection in adult education, refers to *reflective thinking* as an important component of learning. Many other authors have theorized about reflection while offering conceptual frameworks that attempt to clarify the reflection process, its components and outcomes (Boud, Keogh, & Walker, 1985; Mezirow, 1991; Schön, 1983). Known for his work on *transformative learning theory* and *critical reflection*, Mezirow (1978, 1981, 1990, 1991, 1998) and Mezirow & Associates (2000) have been widely quoted in both adult learning and management literatures. Despite this growing interest, Taylor and Cranton (2012) argue that there is a need for more in-depth theoretical analysis on Mezirow's conceptual framework, including new and emerging perspectives. Van Woerkom (2010) advocates that conceptualizations of critical reflection should be grounded in empirical studies. And according to Taylor (1998), many empirical studies assume the presence of critical reflection among participants without actually observing it. Taylor urges researchers to provide "more substantive data of critical reflection and at the same time explore more means for capturing its presence" (Taylor, 2007, p. 186).

Hence, the aim of our paper is to review Mezirow's conceptual work on critical reflection and to identify what we can learn for our own research practice in

HRD from its operationalization in empirical studies. More specifically, we want to answer the following two research questions:

1. *What approaches have been used to operationalize critical reflection?*
2. *How have levels of reflection been assessed, and with what outcomes?*

The term *reflection* can be defined as the activity of exploring or examining an issue of concern and considering it in relation to personal experiences, while *levels of reflection* describe different categories of this activity, often ordered in a hierarchical way. *Critical reflection* depicts the highest category of the “levels of reflection” hierarchy. In our paper, we will refer to either reflection or critical reflection as the activity, and to levels of reflection as the hierarchical categorization scheme of the activity.

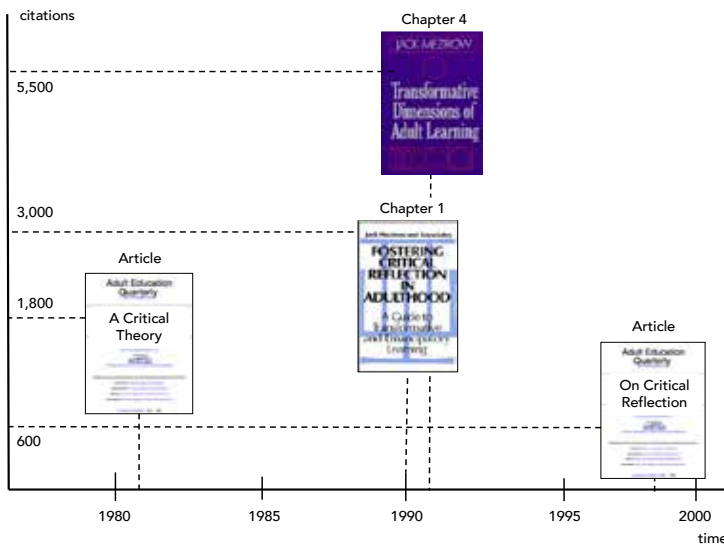


Figure 3.1 Timeline and number of citations of Mezirow’s critical reflection work⁹.

Source (book covers): Mezirow, J. and Associates, *Fostering Critical Reflection in Adulthood: A Guide to Transformative and Emancipatory Learning*, Jossey-Bass Publishers, San Francisco, Copyright © 1990 by John Wiley & Sons, Inc and Mezirow, J., *Transformative Dimensions of Adult Learning*, Jossey-Bass Publishers, San Francisco, Copyright © 1991 by Jossey-Bass, Inc.

⁹ According to Google Scholar, November 2015: 1,997 (1981), 3,300 (1990), 6,360 (1991) and 693 (1998). It should be noted that the number of citations for the 1990 and 1991 book chapters are congruent with the citation of the whole books.

Mezirow on critical reflection

Jack Mezirow, an American sociologist and emeritus professor of adult and continuing education at Teachers College, Columbia University, died on September 24th, 2014. Thirty-six years earlier, when he published his first paper on *perspective transformation* (Mezirow, 1978), only a few would have thought that he would leave a legacy in adult learning to inspire thousands of researchers and practitioners in North America and around the world to think about, study and question the term *critical reflection* as part of the transformative learning process. Reading through his early work again, we noted that “critical reflection” was not mentioned even once in his 1978 publication, but Mezirow made reflection and its levels a significant theme in four publications (see Figure 3.1). By reviewing these often discussed, frequently cited and sometimes contested original publications, we aim to trace the trajectory of Mezirow’s theory on critical reflection, as well as different levels of reflection described in his writings.

Critical consciousness

Mezirow’s (1981) first reflection framework consists of seven levels of reflection. Here, the distinction between consciousness and critical consciousness was first explained by Mezirow. Theoretical reflectivity, i.e., the highest level of *critical consciousness* in this model, depicts the process central to perspective transformation that is the “emancipatory process of *becoming critically aware of how...we see ourselves and our relationships...and acting upon these new understandings*” (Mezirow, 1981, p. 6, *emphasis in original*). Perspective transformation thus describes the major change in how we see the world around us and how we review our deep-seated beliefs about this world through the use of theoretical reflectivity and *critical consciousness*. Mezirow first observed such perspective transformation among women (including his wife Edee) who re-entered college after a long hiatus. In this 1981 paper, references are made to Jürgen Habermas (1971) and two of Mezirow’s collaborators (Broughton, 1977; Maudsley, 1979); no reference is yet made to Dewey (1933) or Dewey’s theory of reflective thinking – a theory that otherwise is often cited as the foundation of critical reflection in adult education.

Non-reflective action

Boud, Keogh and Walker (1985) define reflection as “a generic term for those intellectual and affective activities in which individuals engage to explore their experience in order to lead to new understanding and appreciation” (p. 19). In his 1990 publication, Mezirow comments, “although such a broad definition faithfully reflects common usage, the term needs additional analysis to differentiate reflection from thinking or learning, of which it is a part” (Mezirow, 1990, p. 5). Building on Dewey’s definition of reflection, “assessing the grounds of one’s beliefs” (Dewey, 1933 as cited in Mezirow, 1990, p. 5), Mezirow lays out a second framework of reflective thinking. Here, he distinguishes between *non-reflective action* and reflective action. *Non-reflective action* includes all human action that is habitual or thoughtful without reflection; all other action is reflective. Most people consider introspection as reflective, such as thinking about one’s self, considering one’s thoughts or feelings. However, Mezirow later categorizes introspection as non-reflective because “introspection does not involve validity testing of prior learning” (Mezirow, 1991, p. 107). The term *critical reflection* is introduced as “reflection on presuppositions” (p. 7), and Mezirow (1991) comments that it would be more exact to speak of *premise reflection*. However, he continues to use the more familiar term “as so many of us have used critical reflection to mean the same thing [that] it seems better to continue this practice” (Mezirow, 1991, p. 12).

Content, process, premise reflection

The use of terms changed in his subsequent publication where Mezirow (1991) elaborates in great length on Dewey’s definition of reflection. Clearly, Mezirow now built his reflection terminology on Dewey’s work stating that “reflection is the process of critically assessing the content, process, or premise(s) of our efforts to interpret and give meaning to an experience” (Mezirow, 1991, p. 104). Explanations follow on how content, process and premise reflection differ from one another: *Content reflection* is an examination of the content or description of the problem. This is similar to asking, “*What is happening here? What is the problem?*” (Cranton, 2006b, p. 34, *emphasis in original*). In comparison, *process reflection* focuses more on the problem-solving strategies. “It is asking questions of the form, *How did this come to be?*” (Cranton, 2006b, p. 34, *emphasis in original*). Both content and process reflection could start a reflective process that could lead to emancipatory learning, similar to single-loop learning (Argyris & Schön,

1974). Finally, premise reflection takes place when underlying assumptions or the problem itself are questioned: “*Why is this important to me? Why do I care about this in the first place?*” (Cranton, 2006b, p. 34, *emphasis in original*). Premise reflection involves us in becoming aware of why we think, feel or act in a certain way, and it can be compared to the qualities of double-loop learning (Argyris & Schön, 1974).

Critical reflection of and on assumptions

The foundations of Mezirow’s (1998) *critical reflection of assumptions* can be found in the philosophical work on critical thinking by Harvey Siegel (1988), as well as the empirical evidence presented by P. M. King and Kitchener (1994) who lay out a reflection judgment model that is structured in seven stages. Mezirow’s (1998) framework consists of *objective reframing* on the one end and *subjective reframing* on the other end of the spectrum: “The distinction...is that the former is a consideration of the assumptions, whereas the latter is a consideration on what caused the assumptions to occur” (Kitchenham, 2008, p. 117). For example, objective reframing can be a narrative critical reflection *of* assumptions that requires a critical examination of something that has been communicated to you, while subjective reframing is more the critical self-reflection *on* assumption and what caused the assumption to occur. Mezirow neither refers to Dewey’s reflective thinking in this later article nor does he address non-reflective action or content, process and premise reflection.

Mezirow in HRD

Mezirow’s versatility when it comes to defining critical reflection and levels of reflection over the course of his fragmented high-level conceptual work has been often criticized by contemporary researchers (Kreber, 2012; Kucukaydin & Cranton, 2013; Newman, 2012; Taylor & Cranton, 2012). What is remarkable is that every publication sheds light on a new aspect of reflection: be it the distinction between consciousness and critical consciousness in 1981, the inclusion of non-reflective action in 1990, the explicit reference to Dewey’s work on reflective thinking including content, process and premise reflection in 1991, or the elaboration on critical reflection of assumptions in adult learning in 1998.

Mezirow’s conceptual work on *critical reflection* within transformative learning theory has impacted adult education, HRD and organizational learning litera-

tures. For example, the concept of *double-loop learning* (Argyris & Schön, 1974) is often discussed in conjunction with *premise reflection*, and Schön (1983) builds on the concept of reflection and its production of practice in his writings on the *reflective practitioner*. While Mezirow draws from different intellectual traditions, his focus has been on the individual and how the individual learns through experience and reflection. Mezirow has made an impact on those people who worked with him and who now teach and research in HRD, especially in the adult education side of HRD. These include Marsick and Yorks at Columbia University (see for example Marsick & Maltbia, 2009; Yorks & Marsick, 2000), Bierema (1999) and Watkins (1989) at the University of Georgia, and Cseh (2003) and Scully-Russ (2005) at George Washington University.

Methodology: Conducting a step-by-step literature review

In order to select and review empirical studies that operationalize reflection and its levels, we conducted a literature review along a structured step-by-step process:

1. Defining selection criteria
2. Devising a search strategy
3. Conducting the literature search
4. Reviewing selected studies

First, we formulated selection criteria. In order to be included in this literature review, publications would need to (1) build on Mezirow's critical reflection definition, (2) represent an empirical study with a methodology section, (3) be published in a peer-reviewed journal (4) and be written in English language. A study would be excluded if it would (1) be rooted in a different tradition of critical reflection, (2) represent a purely conceptual or theorizing body of work, or (3) did not provide sufficient supplementary information on data collection and analysis, such as the questionnaire or coding scheme used. This last exclusion criterion was especially important since our aim is to retrace the opera-

tional definition of critical reflection in empirical studies and what we can learn from it for our own research in HRD.

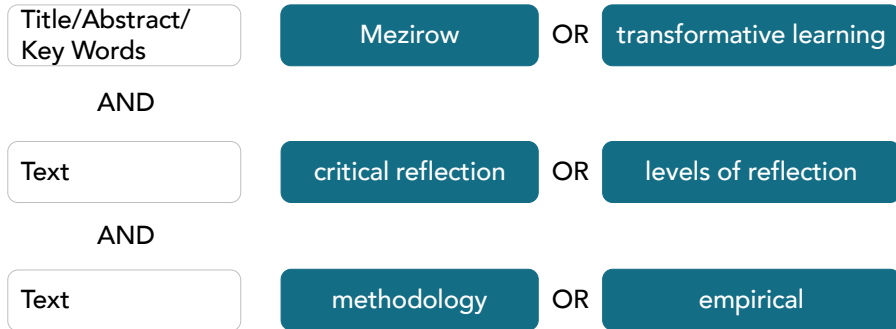


Figure 3.2 Search terms used for literature review

Second, we devised a search strategy that was based on previous literature reviews of empirical studies in transformative learning theory, inspired by earlier reviews conducted by Taylor (1997, 1998, 2007) and Snyder (2008). A number of key publications were analyzed to identify relevant search terms. We used “snowballing” (Callahan, 2010) to identify references in articles that were found in Taylor’s and Snyder’s reviews, and we also used reverse citation tracking to find articles that cited studies already deemed relevant to our review. We found that the term *reflection* was used in many different forms, e.g., *critical reflection*, *reflective thinking/learning/practice*, *levels of reflection*, etc. Some of the researchers were interested in the empirical examination of an existing coding scheme, while others were focused on deriving a new coding scheme or a mix of both. After testing different search terms, we found that a specific combination yielded high-quality results suitable for our literature review (see Figure 3.2).

Third, a literature search was performed in three databases in September 2014: EBSCOhost (incl. PsycINFO, ERIC, Socindex, Medline and Econlit), ProQuest and PubMed. After removing duplicates, the search resulted in 213 unique publications that we screened along the inclusion/exclusion criteria. A large number of studies were excluded after screening the abstracts as they did not focus on Mezirow’s definition of critical reflection or levels of reflection. In fact, 37

studies were selected for a full-text review in which careful consideration was given to the methodology used and the supplementary information on data collection and analysis provided. The lack of sufficient information on how the reflection framework was operationalized during the research process was the main reason for excluding a study at this final stage of our review. We ended up retrieving 12 publications that were included in this literature review. See Figure 3.3 for more details on the review process, steps and selection outcome.

Finally, the selected studies (N=12) were analyzed and compared using a tailored review form. For each study, background information such as research setting, design, country, target group and sample size were recorded, as well as the theoretical reflection framework that the study was based on. When reflection was coded, different criteria to assess the level of reflection were used. For example, some researchers analyzed individual paragraphs of the interview transcripts; others looked at the interview transcript or reflection journal as a whole. To compare these different assessment criteria, we first looked the underlying theoretical frameworks and then the assessment process of each study, as well as the actual reflection outcomes. See Tables 3.1 and 3.2 for an overview of the publications reviewed.

In sum, the selected studies were published between 1995 and 2011 in various educational journals. The *Journal of Transformative Education* was the place of publication of three of the selected articles, which does not seem too surprising given the specific search terms. All but one of the studies listed in their appendices the instruments or coding schemes used. Where additional information was not displayed, we contacted the author(s) to retrieve the relevant questionnaire(s).

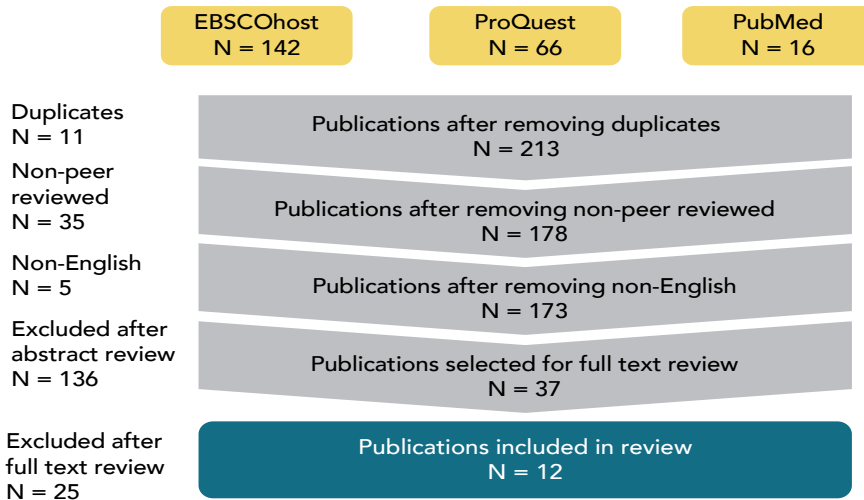


Figure 3.3 Literature review process steps and outcome

Approaches to operationalize critical reflection

In this section of the paper, we analyze a variety of research approaches that are used in operationalizing critical reflection. We will refer to categories of comparison used in literature reviews common in education research (Newton, Rothlingova, Gutteridge, LeMarchand, & Raphael, 2011). These include study aim, research design, data-collection method and setting (see Table 3.1).

Table 3.1 Summary of approaches to operationalize critical reflection

Authors (year)	Study's aim	Design	Data collection	Setting	Country
Richardson & Maltby (1995)	To analyze reflection and learning through reflective diaries in nursing education	Cross-sectional	Journal entries and focus groups	Undergraduate nursing students (N=30 journals and N=8 groups)	Australia
Kember et al. (1999)	To develop a coding scheme to estimate the quality of reflective thinking in reflection journals	Instrument development	Journal entries and interviews	Undergraduate students in health sciences (N=9 for practical test)	Hong Kong
Kember et al. (2000)	To develop a questionnaire to measure the level of reflective thinking	Instrument development	Questionnaire	Undergraduate and postgraduate students in health sciences (N=303)	Hong Kong
Liimatainen et al. (2001)	To analyze levels of reflectivity of student nurses in counseling situations	Longitudinal, 3 interviews over 3 years	Interviews (incl. videos to stimulate recall)	Nursing students (N=16)	Finland
Cranton & Carusetta (2004)	To develop a model on authenticity in higher education teaching	Longitudinal, 8 interviews over 2 years and focus groups in year 3 for model development	Interviews, observations and focus groups	Educators/faculty in higher education (N=22)	Canada
Kreber (2005)	To test a reflection framework in the context of academic staff development	Cross-sectional	Interviews and questionnaire	Science instructors in higher education (N=36)	Canada
Peltier et al. (2005)	To apply an instrument that measures reflective learning	Cross-sectional	Questionnaire	Graduate students post-completion of MBA program (N=220)	USA

Authors (year)	Study's aim	Design	Data collection	Setting	Country
Kitchenham (2006)	To investigate the educational technology development of teachers through the lens of transformative learning theory	Cross-sectional	Journal entries, interviews and researcher's field notes	Elementary school teachers (N=10)	Canada
Chirema (2007)	To analyze reflective journals using two models	Cross-sectional	Journal entries and interviews	Part-time post-registration students in nursing (N=42 journals; N=20 interviews)	UK
Wallman et al. (2008)	To test a coding scheme for assessing students' levels of reflection	Cross-sectional	Journal entries (essays)	Pharmacy internship students (N=126)	Sweden
Kitchenham & Casteauneuf (2009)	To explore the appropriateness of a specific framework to examine critical reflection in e-portfolios	Cross-sectional	Journal entries (rationale statements)	Post-degree program students in teacher education (N=127)	Canada
Bell et al. (2011)	To evaluate the effectiveness of a specific coding scheme to assess reflection in written journals	Cross-sectional	Journal entries	Undergraduate students in business education (N=7)	Australia

Study aims

The 12 studies can be broadly categorized into three types of aims. The first type of studies looks into developing a new instrument or conceptual model to study and assess reflection. In these studies, the aim usually started with “to develop a coding scheme / questionnaire / model that...” (Cranton & Carusetta, 2004; Kember et al., 1999, 2000). The creation of something new is a main driver for this type of research study. Another set of studies looked at testing or evaluating an existing coding scheme: “To test / evaluate a...” is the common wording used with this second type of study (Bell, Kelton, McDonagh, Mladenovic, & Morrison, 2011; Kreber, 2005; Wallman, Lindblad, Hall, Lundmark, & Ring, 2008). The third group of studies falls somewhat in between; here the aim was to look at an existing scheme and then to add to it in order to adapt it to the purpose of the specific study. Study aims in this “in between” group are described with the verbs “to analyze / to apply / to investigate / to explore” (Chirema, 2007; Kitchenham, 2006; Kitchenham & Chasteauneuf, 2009; Liimatainen, Poskiparta, Karhila, & Sjögren, 2001; Peltier, Hay, & Drago, 2005; Richardson & Maltby, 1995).

These different types of aims can be paralleled with a taxonomy named after Benjamin Bloom (1956), the educational psychologist who developed a classification of levels of intellectual behavior that are relevant in learning. Applying a revised version of Bloom’s Taxonomy (Krathwohl, 2002), the first set of studies would fall under the header of “creating”; the second group under the header of “applying” and the third group somewhere between, into “analyzing” or “evaluating.”

Research designs

In terms of research approaches, quantitative and qualitative approaches could be observed or a combination of both. Eight studies followed a cross-sectional design and two involved a longitudinal research design. The longitudinal approach was hence much rarer; this is understandable, as a research study that follows the same participants over a period of time is much more time- and cost-consuming than an investigation that takes place at one point in time in the cross-sectional approach. Liimatainen et al. (2001) reason that their longitudinal design was worthwhile notwithstanding the smaller number of participants (N=16), as they “managed to follow the development of reflective practitioners

in the area of health promotion during a 3-year nursing education program” (p. 657). The remaining two studies both led by David Kember and colleagues were also cross-sectional in nature but with the intent to develop a research instrument. Kember et al. (1999) analyzed conceptual studies and frameworks on the topic but “felt that articles in the existing literature did not meet the criteria we specified for a method of assessing the level of reflective thinking” (p. 19). The authors therefore suggested that it was justifiable to develop a new scheme based on existing conceptual work that would fulfill their need.

Data collection

Seven out of 12 studies based their findings on written accounts, e.g., journal entries, diaries, reflection essays or rationale statements. Some of these studies also made use of a complementary data collection method, such as interviews or a questionnaire. Three studies based their analyses on personal accounts collected during interviews, some of which were followed up by a focus group or a questionnaire. Whereas studies relying on interviews might be easier to conduct, they are also “limited by the extent to which participants are willing to accurately represent their experience or can correctly remember it” (Kreber, 2012, p. 333). On the other hand, engaging students in reflective journal writing – especially when the research participation is voluntary – also comes with its challenges: “A small number stated that although they had reflected they did not believe it was necessary to write it down” (Chirema, 2007, p. 199). Two studies used the survey method by employing a questionnaire to collect data. The sampling method varied across studies. While Cranton & Carussetta (2004) explain that, “the main criterion for selection was a deep interest in teaching and learning” (p. 277), Richardson & Maltby (1995) ran with “a random selection of 30 diaries from those students who had provided consent” (p. 236).

Table 3.2 Summary of assessing levels and reflection outcomes

Authors (year)	Theoretical framework	Assessment process (units)	Assessment process (themes)	Reflection outcomes	Suggested reflection categories (new categories in bold)
Richardson & Maltby (1995)	Powell (1989) based on Mezirow (1981)	By paragraph By student	No	Low (94% of the total number of scores in lower reflectivity)	Six levels (hierarchical): Reflectivity, affective reflectivity, discriminant reflectivity, judgmental reflectivity, conceptual reflectivity, theoretical reflectivity
Kember et al. (1999)	Mezirow (1994)	By paragraph By student	No	NA	Seven levels: habitual action, thoughtful action, introspection, content reflection, process reflection, content + process reflection , premise reflection
Kember et al. (2000)	Kember et al. (1999) based on Mezirow (1991)	NA (questionnaire)	Yes (health science)	NA	Four levels: habitual action, understanding , reflection, critical reflection as questionnaire items for students in health sciences
Liimatainen et al. (2001)	Mezirow (1984)	By paragraph	No	High (50% of students achieved critical consciousness levels 5-7)	Eight levels: Non-reflective thoughtful action , reflectivity, affective reflectivity, discriminant reflectivity, judgmental reflectivity, conceptual reflectivity, psychic reflectivity, theoretical reflection
Cranton & Carusetta (2004)	Grounded Theory; some reference to Mezirow (1991)	By paragraph	Yes (authenticity)	NA	Four categories of critical reflection: self, others, relationship and context/environment
Kreber (2005)	Kreber (1999) and Kreber & Cranton (2000) based on Mezirow (1990)	By paragraph	Yes (knowledge domains in teaching)	Low (more “declarations of reflection” than concrete indicators found)	3 x 3 model with three levels of reflection (content, process, premise) across three knowledge domains in teaching

Authors (year)	Theoretical framework	Assessment process (units)	Assessment process (themes)	Reflection outcomes	Suggested reflection categories (new categories in bold)
Peltier et al. (2005)	Kember et al. (2000)	NA (questionnaire)	Yes (interaction conditions)	NA	Four levels (habitual action, understanding, reflection and intensive reflection) and two conditions (instructor-to-student and student-to-student) leading to learning outcomes
Kitchenham (2006)	Mezirow (1990, 1995)	By paragraph	No	Moderate (55% of comments coded as premise reflection)	Three levels: content, process, premise (objective reframing; subjective reframing)
Chirema (2007)	Boud et al. (1985) in step 1; Mezirow (1990, 1991) in step 2	By paragraph By student	No	Moderate (two thirds of students showed varying levels of reflection)	Six categories: attending to feelings , association, integration, validation, appropriation, outcome of reflection
Wallman et al. (2008)	Kember et al. (1999) based on Mezirow (1991)	By paragraph By student	No	Low (25% reflection at start and 40% reflection at end of semester)	Six levels: habitual action, thoughtful action, introspection, content reflection, process reflection, premise reflection
Kitchenham & Casteauneuf (2009)	Kitchenham (2008) based on Mezirow (1998)	By paragraph (statement)	Yes (teacher education standards)	NA	Six categories consisting of 2 x objective reframing (narrative and action) and 4 x subjective reframing (narrative, systemic, therapeutic and epistemic)
Bell et al. (2011)	Kember et al. (1999) based on Mezirow (1991)	By paragraph	No	Low (65% non-reflection)	Six levels: thoughtful action, introspection, content reflection, process reflection - others , process reflection - internal , premise reflection

Settings

The studies were conducted in different learning environments: half of them in medical education including nursing, pharmacy and health sciences, and the other half in teacher, business, elementary and higher education or professional development. In terms of geographic spread, the studies took place in industrialized countries across the world, including two studies from Hong Kong. Remarkably, four of the 12 studies were conducted in the Canadian context, which shows the density of transformative learning and critical reflection research taking place in Canada.

Observing a variety of *aims, research designs* and *data collection methods* in a number of different *settings* leads us to the initial conclusion that there is little agreement on the operationalization approach of reflection in adult education and HRD research. To deepen our understanding in this area, we will next explore how levels of reflection have been assessed in these various operationalization approaches.

Assessing levels and reflection outcomes

To answer the second research question, it was harder to find a descriptive framework with set criteria for comparison. Hence, we draw from different sources when reviewing theoretical frameworks (Dyment & O'Connell, 2011), the levels of reflection assessment process (Koole et al., 2011) and actual reflection outcomes (Rogers, 2001). We will first look at the underlying theoretical frameworks and their translation before analyzing the assessment process on levels of reflection employed in each study. Lastly, we depict reflection outcomes and discuss how researchers have critically reflected on their findings (or not).

Translating theoretical frameworks

Table 3.2 shows which studies based their analysis on which of Mezirow's conceptual models.

- Mezirow (1981): Liimatainen et al. (2001); Richardson & Maltby (1995)*¹⁰

¹⁰ The * indicates each study that draws from several conceptual models, Mezirow's framework on critical reflection being one of them.

- Mezirow (1990): Kitchenham (2006); Kreber (2005)*; Chirema (2007)*
- Mezirow (1991): Kember et al. (1999); Cranton and Carussetta (2004)*; Chirema (2007)*
- Mezirow (1998): Kitchenham & Chasteauneuf (2009)*

The frequency of model adoption seems to be in line with the overall “popularity” of each conceptual model (based on citation numbers), which shows the 1990 and 1991 frameworks as more popular (see Figure 3.1).

Four of the 12 studies were based on an instrument developed by Kember et al. (1999, 2000), which had evolved as a model from Mezirow’s (1991) levels of reflection. All of these four studies focus on the assessment of student reflection from journal entries:

- Kember et al. (1999): Kember et al. (2000); Wallman et al. (2008); Bell et al. (2011)
- Kember et al. (2000): Peltier et al. (2005)

All authors who chose to use one of the Kember et al. instruments were working with student populations, and their study aims centered on testing or evaluating the validity, reliability and/or usefulness of these instruments. It seems that both the coding scheme and the questionnaire established by Kember et al. have gained popularity in recent years as they seem to be more frequently used in both business and health education studies.

The theoretical framework selection seems to be closely linked to the studies’ aims. Wallman et al. (2008), who were looking to test an existing coding scheme, stated that they had selected Kember et al.’s (1999) scheme as it had a sound theoretical foundation and it had been empirically tested before with satisfactory results. Similarly, Peltier et al. (2005), who were looking to apply an instrument with some modifications, reasoned their choice as follows: “When developing our questionnaire, we did an extensive review of the reflection literature...The work of Kember et al. (2000) was especially helpful” (p. 257).

However, some major obstacles had to be overcome when translating the theoretical framework into some coding scheme that had not been empirically tested before. Kember et al. (1999), who did the groundwork in the assessment of reflection levels in student journals, describe what additional data were collected to aid the coding and interpretation of journal data: "In an attempt to clarify the intention of the student in this piece of journal writing, the student was interviewed by one of the research team a few weeks after the journal was written" (p. 27). In a similar way, Richardson & Maltby (1995) decided to conduct focus groups, and "the process of writing up the tape-recorded interview proved crucial in eliciting rich, detailed material that could be used to support and explicate the meaning of the tabulated findings" (p. 239). Hence, a number of studies made use of multiple data collection pathways in order to enhance their understanding and facilitate the assessment of reflection levels.

Selecting units in assessment process

Whether the study used authored documents in the form of reflective journals or student diaries, or whether the data collection focused on verbatim transcripts of interviews or focus group sessions, in each case, researchers had to decide on a coding scheme and a coding unit. The coding schemes were often either the instrument itself or a variation of it that was adjusted to the specific context of the study. Approximately half the researchers decided to analyze the data in broken-down units, i.e., by paragraph in order to establish the level of reflection for each coded unit. About one-third of the researchers coded written documents paragraph-by-paragraph or section-by-section, but also established a summative level of reflection per student. The advantage of this two-step coding process is to have countable coded paragraphs, as well as an assessment of the student's overall level of reflection. Wallman et al. (2008) emphasize this two-step process where a coder "read all essays one time, categorized sections in a second reading, and the whole essay in a third reading, while controlling for possible errors" (p. 5). Other than two studies using questionnaires that were analyzed through statistical modeling (Kember et al., 2000; Peltier et al., 2005), no further variation of coding units was used.

Opting for themes in the assessment process

Another differentiator was the inclusion of reflection subjects or themes into the coding scheme and assessment process. Most coding schemes that are based on

Mezirow's conceptual work employ some form of hierarchical levels, e.g., content and process reflection of "lower order" reflection compared to premise reflection. However, these hierarchical models on levels of reflection do not give away the subject or theme of reflection. This might not be an issue, as some researchers are less interested in the content of reflection, but more so in the assessment of the reflection level. For other authors, however, the context-embeddedness of their coding schemes plays a bigger role, and in fact the reflection themes employed are integral to their research. In Kreber's (2005) reflection framework, however, reflection themes were included. Teachers could, for example, reflect on some form of pedagogical knowledge by asking themselves: "What do I know about how students learn?" which would be a separate theme from reflecting on the instructional design of a lesson. Kreber calls these themes "knowledge domains", and she integrates them into her coding scheme by pairing each theme with each reflection level, establishing a 3 x 3 coding matrix.

Five of the 12 studies presented in this literature review make use of this quality of embedding reflection themes into their coding schemes and assessment processes (Cranton & Carusetta, 2004; Kember et al., 2000; Kitchenham & Chasteauneuf, 2009; Kreber, 2005; Peltier et al., 2005); the remaining seven studies look at levels of reflection without referring to reflection themes.

There is one more interesting finding in relation to coding. Kreber (2005) reports in her study on science instructors that there seems to be a large difference between the *declarations* of reflection in the interview setting in comparison to finding *concrete indicators* of reflection. This means that what a person states is not necessarily what this person actually does. Trying to explain this phenomenon, Kreber (2005) argues:

It is possible that they really do engage in reflection but do not know how to show it. It is equally possible that they think they engage in reflection, but they do not really do it because they do not know how to (p. 352).

This is an important finding as it highlights the limitations of inquiry on the subject of reflection regardless of the coding scheme, instrument or coding unit.

Assessing reflection outcomes

Whether or not the highest level of reflection was achieved is often the reported focus in the results section of a study on critical reflection. Here, the levels of reflection might either be described as “high”, defined in this review as 50% or more people in the study achieving the highest level, or as “low”, where 10% or less achieved the highest level of reflection or where authors stated it as low. Any reflection outcomes that were reported between high (>50%) and low (<10%) we have here classified as “moderate”:

- High: Liimatainen et al. (2001)
- Moderate: Kitchenham (2006), Chirema (2007)
- Low: Richardson & Maltby (1995), Kreber (2005), Wallman et al. (2008), Bell et al. (2011)

The remaining studies did not establish levels of reflection as an outcome; this is especially true for those studies that focused on developing a new instrument or coding scheme.

It can be observed that high frequencies of critical reflection occur relatively seldom. This is in line with the common understanding of critical reflection and the potential for a *perspective transformation*, i.e., that people experience a major change of perspective along an alteration of their deep-seated beliefs. This perspective transformation does not happen each and every day, neither is all learning intended to yield transformation as an outcome: “Why should a student have to change as a person in order to get a high grade?” (Bell et al., 2011, p. 809).

This concern with reflection outcome “fixation” is shared with more researchers who discuss the assessment of reflection levels of students in an educational setting. For example, Wallman et al. (2008) question: “Did the scheme really measure reflection or was it the students’ ability to express thoughts in written form that was assessed?” (p. 6). While some researchers display no concerns in using reflection schemes for summative assessment (Kember et al., 1999, 2000; Liimatainen, Poskiparta, Karhila, & Sjögren, 2001), others prefer to regard such reflection schemes and categories as formative tools to encourage dialogue and the development of reflective thinking skills (Bell et al., 2011; Peltier et al., 2005; Richardson & Maltby, 1995; Wallman et al., 2008).

Notwithstanding the conclusion drawn at the end of each study, it can be stated that researchers spend a lot of time on coding and assessing reflection, with often ambivalent results. “Why critical reflection was not frequently observed is a question that is of interest,” states Kreber (2005) in her concluding remarks. She speculates whether academics in a scientific discipline would exhibit higher levels of critical reflection as compared with the levels found among social scientists and academics in the humanities.

Suggesting new reflection categories

Going forward, the 12 studies suggest multiple models on critical reflection, with the number of reflection levels ranging between three and eight. Reviewing the last column in Table 3.2, it can be observed that the reflection models keep expanding, e.g., when Kember et al. (1999) added the “content and process reflection” level to Mezirow’s (1991) conceptual scheme or when “critical reflection” becomes “intensive” reflection in Peltier et al.’s (2005) suggested model. Further additions of categories include “non-reflective thoughtful action” (Liimatainen, Poskiparta, Karhila, & Sjögren, 2001), “instructor-to-student” and “student-to-student” conditions for reflection (Peltier et al., 2005), and the “attending to feelings” category suggested by Chirema (2007). Some of these suggested new reflection categories make meaningful additions to the current reflection concepts. For example, the “attending to feelings” category proposed by Chirema would complement the more cognitive components of the reflection process. Furthermore, including “instructor-to-student” or “student-to-student” conditions for reflection in the questionnaire or coding scheme as modeled by Peltier et al. (2005) would give the reflection process a social dimension that is relevant for understanding learning in any type of institutional or organizational context.

Four ways of improving critical reflection research

Our starting point was Mezirow’s conceptual work on critical reflection and what we could learn from it: How can we improve our own critical reflection research in HRD by comparing its operationalization in empirical studies? We conducted a literature review showing that Mezirow’s theory has been translated using different approaches, assessment processes and outcomes in research studies. The review also confirms that little progress has been made when it

comes to the integration of different reflection models into harmonized instruments that can be applied in different disciplines of the adult education and HRD spectrum. We discussed the studies' limitations, and propose with four ways of improving critical reflection research going forward.

Integrating different critical reflection traditions

Reflection has been theorized, researched and reviewed in different ways, across different disciplines and from different viewpoints. According to Brookfield (2000) critical reflection research has been shaped by different, and often conflicting, intellectual traditions that inform the use of the term "critical reflection". Van Woerkom (2010, 2008) summarizes these six traditions, which include (1) the tradition of ideology critique – think: the Frankfurt School and the development of critical theory between 1930 and 1970; (2) the psychotherapeutically inclined tradition, with a focus on the individual, with human experiences from childhood onward, and self-examination of feelings and thoughts to develop new perspectives; (3) the tradition of analytic philosophy and logic, e.g., represented in the work of Ennis (1993), whose work emphasizes the importance of knowing what to believe or do; (4) the tradition of pragmatist constructivism, elements of which are evident in Dewey's (1933) work; (5) the tradition of qualitative social science, which emphasizes the role of the researcher as a reflective practitioner; and (6) the tradition of organizational learning that relates to concepts such as Argyris and Schön's (1996) double-loop learning.

Different perspectives influenced Mezirow, most notably the psychotherapeutically inclined tradition on which we focused the attention of our literature review. However, Mezirow also rooted his thinking in the tradition of ideology critique (think: Jürgen Habermas) as well as pragmatist constructivism, where Mezirow drew heavily from Dewey to explore the role of experience and reflection in adult learning. Furthermore, Mezirow got inspired by Paulo Freire who led the critical pedagogical movement. Although our focus in this literature review has been on the individual with the aim to better understand critical reflection and learning at this psychological level, we believe that the same literature review with a different focus could enrich our thinking on reflection and its outcomes. For example, an exploration in the tradition of organizational learning (Argyris & Schön, 1996; London & Sessa, 2006; Antonacopoulou & Sheaffer, 2014; Cooper, 2014) that takes into account the individual *and* the orga-

nizational context would make an interesting addition to our review. We hence suggest replication of this study with a different critical reflection tradition as the backdrop for future research.

Using multiple data-collection pathways

It becomes evident that reflection can be operationalized using multiple approaches and different pathways: from cross-sectional studies to longitudinal designs and from the collection of written accounts to the transcription of interview materials. As we showed in our analysis, some studies focus solely on written material, while other studies narrow in on verbal communication that was collected during interviews and focus groups. Some studies integrate written and verbal data collection methods.

The question arises whether this has to be an “either-or” choice, or whether there might be the opportunity to mandate a combination of multiple data collection pathways in this line of research. Social researchers generally argue that using different sources helps to elicit rich, detailed material – a statement that is definitely true for research studies on critical reflection. Less uncertainty during the coding and interpretation process and more interaction with the research participants seem to be positive outcomes of choosing not a single but multiple data-collection opportunities. Study results might become sharper through increased and enriched communication between researcher team and participants. Gray (2007) explores the practice of critical reflection within a management learning process and concludes that there are different reflection tools that can be applied in a managerial setting, including reflective journals. Nesbit (2012), however, argues that the use of a reflective journal may present as an unlikely activity for a busy leader who already struggles to find time to carry out reflective analysis of events experienced, let alone engage in a disciplined approach to journal writing (p. 212).

In yet a different setting, researchers added observation to their study of critically reflective practice in a counseling organization (Keevers & Treleaven, 2011). Hence, how to capture the moment of reflection in the workplace may differ in each professional setting.

Triangulation of observation, written and verbal accounts increases the chance that our coding efforts as researchers actually result in meaning-making. Not only will this approach of using multiple data collection pathways help participants to address, express and articulate reflection in different ways, but it will also enable researchers to understand more of the individual's social environment in which his or her reflection takes place. The underlying assumptions of this literature review were that reflection could, indeed, be articulated and that reflection is, in fact, documentable. Here, a neuroscientist might argue that functional magnetic resonance imaging (fMRI) of the brain would be an alternative and maybe more precise way of capturing a person's level of reflection (Berman, Jonides, & Nee, 2006). We acknowledge this potential flaw in our social science research methodology, which depends on an individual's self-awareness, articulation and judgement. However, even if an fMRI scan could give us reliable data on a person's critical reflection in the future, would we be able to interpret its meaning and relevance without actually talking to the individual? To bridge this gap, we suggest collecting concrete indicators of reflection in addition to declarations of reflection using multiple data collection pathways.

Opting for thematic embedding

The analysis further shows that reflection levels are translated in two forms: either in the "pure" form or using contextual information relating to the research setting that marries reflection levels with relevant reflection themes. In our view, this situating of reflection levels is an important building block of any coding scheme in critical reflection because reflection happens on a subject that matters to the person who reflects. Ignoring the reflection theme or the circumstance of reflection would reduce reflection to a thematically "empty" exercise.

It should be noted that ten of the 12 studies reviewed for this analysis were conducted with students. The other two studies were with faculty and teachers and not with the working populations that are typically the focus of HRD research. As such, can the findings related to situating reflection be transferred from studies with students to a working population? We think that this inference can be made, especially as we found several HRD studies of working populations that discuss reflection themes and differences between these. For example, Berings et al. (2008) have found a distinction between reflection on knowledge, on skills and on attitude in their study of on-the-job learning within the nursing

profession. Nesbit (2012) conducted a conceptual study on self-reflection in leadership development. Here, the author calls out for more qualitative studies in different development contexts, “such as expatriate assignments, new roles, ... as well as in different industry and strategy contexts” (Nesbit, 2012, p. 218) to gain greater understanding of reflection processes of managers and leaders.

Depending on the area of HRD and workplace learning, we should be able to amend the reflection themes, and hence elaborate on the coding scheme or instrument in its specific setting. For example, if we look at levels of reflection in HRD initiatives at a hospital, the themes could embrace different knowledge domains of patient care; for a professional development program in the construction industry, we could think about knowledge domains in building and maintenance; and for a study in business administration the domains could be split up into knowledge domains, such as financial acumen, leadership and time management. Whether to opt for thematic embedding depends on the aim of the research. As a consequence, thematic embedding would raise local and theme-specific validity, but it does not make it easier to compare various settings and it might potentially be more difficult to come to an overriding view.

This finding on thematic embedding might call for a more *situative perspective* (Fenwick, 2000a), in which learning is tacit and can be understood by observing people and how they interact with their environment to learn and generate knowledge. Examples of more situated studies of reflection and learning in the workplace can be found in Lundin & Nuldén’s (2007) study of Swedish police officers and Jordan’s (2010) research on the professional practice of anesthesiologists in Central Europe. In our literature review, we had limited ourselves to studies that focus on the individual’s perspective and that draw from self-reported data. Adopting a more situated approach that researches the individual, the environment, and their interaction could provide another fruitful ground for future study.

Attending to feelings

Finally, a lot of attention, time and energy have been spent on measuring reflection, its levels and its outcomes. We showed this in the later part of our analysis. The focus in adult education and HRD research should shift away from reflection outcomes and toward the reflection process and its boundaries with

non-reflective activities. When does thinking lead into reflection, and what triggers reflection to become critical reflection? We acknowledge that there is more research available on triggering events, e.g., in the form of disorienting dilemmas (Mälkki, 2012) or critical incidents (Cope & Watts, 2000), and that this research would complement our analysis.

Another question is whether critical reflection research should aim to capture *non-reflective* activities, such as habitual or thoughtful action, or anything that relates to emotion rather than cognition. This brings us to the “rationalistic bias” (Van Woerkom, 2010, p. 347) that underlies almost all reflection research: We try to separate emotion from cognition and aim to distill pure thinking without dipping into any extra-rational aspects of experience and emotional learning. Callahan (2004) advocates to break the “cult of rationality” by raising awareness of the implications of emotions for adult education and HRD. We therefore suggest broadening our cognitive perspective on the boundaries of critical reflection to allow for non-reflective and emotional aspects, too, as other researchers have already indicated (Cranton, 2006a; Dirkx, Mezirow, & Cranton, 2006; Tisdell & Tolliver, 2001).

For example, Mälkki (2010) makes a link between Mezirow’s and Damasio’s theories where Mälkki states that “the challenges of reflection are fundamentally connected to the way the biological life-support system affects our thinking through the emotions” (p. 58). Also, Dirkx (2008) looks at the emotional side of adult learning, and he promotes that emotion should play a more integral, central and holistic role in reason, rationality, learning and meaning-making. Furthermore, Yorks & Kasl (2002) challenge the rationalistic view by exploring whole-person learning when taking into account experience and the role of affect.

We found that only a few authors have incorporated emotional aspects of learning, for example the work by Cope & Watts (2000), who research the learning processes of entrepreneurs by using the critical incidents technique. The researchers asked about critical moments in the history of the business as a departing point to explore personal development and organizational growth. Cope & Watts (2000) state that emotion-laden experiences are often the basis of these critical incidents, and that these experiences lead to personal development,

learning and growth. Also, Callahan & McCollum (2002) have elaborated on this topic by conceptualizing emotion research in the organizational context. More research that investigates feelings, affects or emotional aspects of experience and learning, possibly narrated through critical incidents, would be welcome.

Checklist for future research

We conclude that research in this field seems to be obstructed by the lack of an overarching framework for how to operationalize critical reflection. We acknowledge the main limitations in our literature review: our restricting focus on one tradition of critical reflection, the reliance on self-report data, the transferability of our findings into other contexts, and the omission of emotional responses in the studies analyzed. We suggest the following checklist for future research on critical reflection:

1. When establishing a conceptual reflection framework, have I looked at different traditions of critical reflection research?
2. When setting up the research design, am I using multiple data collection pathways to record, capture and collect meaning-structures of my participants' reflection processes and outcomes?
3. When stimulating reflection recall during data collection, have I embedded my questions in study-relevant themes?
4. And finally, how am I attending to participants' feelings in the overall reflection research process?

Chapter 4

“This is not a test”

How do HRD professionals use personality tests as tools of their professional practice?¹¹

¹¹ This chapter has been published as: Lundgren, H., Poell, R. F., & Kroon, B. (2019). “This is not a test”: How do HRD professionals use personality tests as tools of their professional practice? *Human Resource Development Quarterly*.



Abstract

Although Human Resource Development (HRD) professionals enjoy the use of personality tests in their practice, the appeal of these tests to some is harshly criticized by others. Personality tests attract through optimistic descriptions and ease-of-use for individual and team development while often lacking predictive and discriminant validities. Despite those concerns, the personality testing market can be characterized as a dynamic industry, with many professionals using assessments in developmental settings such as management training and executive coaching. The aim of this paper is to explore how individual meaning-making and organizational sensemaking theories help to explain the widespread and sustained use of personality tests in developmental contexts among HRD professionals. Using grounded theory and inductive analysis, we distill meaning from semi-structured interviews with 18 HRD professionals. Through pattern analysis, we establish six strategies that describe practical approaches in personality testing: 1. *Ethical-protective*, 2. *Scientific-selective*, 3. *Cautious-avoiding*, 4. *Cautious-embracing*, 5. *User friendly-pragmatic* and 6. *Knowledgeable-accommodating*. We find that HRD professionals deal with cognitive dissonances and paradoxical situations in their professional personality test use practice on a regular basis. Research limitations and implications for practice and future research are discussed.

Introduction

Survey research conducted by the Chartered Institute of Personnel and Development (CIPD) in the UK shows that personality tests are increasingly employed by Human Resource Development (HRD) professionals for individual and team learning purposes (McGurk and Belliveau, 2012). Personality tests are standardized assessments used to determine a person’s set of preferences, traits or behavioral styles, most commonly – but not exclusively – using self-report questionnaires. Human Resource (HR) practitioners may use these tests for recruiting (Hossiep et al., 2015) as well as for development purposes (Benit & Soellner, 2013), although many of the personality tests used were originally developed for personnel selection and not specifically for developmental practice (McAdams, 1997; McCrae & McCrae, 1996).

When looking at research in the developmental context more closely, we find that personality tests are used in management coaching (Nelson & Hogan, 2009; Passmore, 2008; Passmore et al., 2010; Scharlau, 2004), educational leadership (Tomlinson, 2004), organizational and team development (Badham et al., 2003; Clinebell & Stecher, 2003; Kuipers et al., 2009; Ludeman, 1995), management learning (Ford & Harding, 2007; Furnham & Jackson, 2011; Goodstein & Prien, 2006), and leadership development (Allen & Hartman, (2008). In many of these studies, type-based personality tests, and in particular the Myers-Briggs Type Indicator (MBTI)¹², are named as the most popular tools for developmental use (Clinebell & Stecher, 2003; Passmore et al., 2010; Stothart, 2011). A type-based personality test may appeal to HRD professionals as the descriptions of personality are mostly optimistic (Ford & Harding, 2007), feedback reports ease the exploration of differences among test takers (Passmore et al., 2010) and discussion of personality test outcomes promises to facilitate team development (Clinebell & Stecher, 2003). However, the appeal of these tests to some is harshly criticized by others: Type-based tests generally have poor predictive validity (Furnham & Crump, 2005; Gulliford, 1991; Gardner & Martinko, 1996; Pittenger, 2005), low discriminant validity (Fisher, Hunter, & Macrosson, 2001) and methodological issues due to forced-choice answer formats (Converse et al., 2008; Harland, 2003). This leads to a paradoxical situation (Lewis, 2000; Lewis

¹² For a list of personality test acronyms and their descriptions, see Appendix 1.

& Smith, 2014) wherein personality tests are popular among HRD professionals but at the same time are one of the most critiqued instruments, too (Boyle, 1995; Pittenger, 2005). It is unclear how HRD practitioners deal with the tensions that arise from this paradox. In order to inquire more deeply into the practice of HRD professionals who apply personality tests in developmental settings, we will first elaborate on the test industry before looking at studies that have analyzed HR practitioner reactions to tests.

Test industry data shows that the personality testing industry overall is an expanding market: With about 2,500 personality tests administered a few million times every year, they generate approximately \$500 million per annum in test license and certification revenue for test publishers in the USA (Weber and Dvoskin, 2014). It is also a dynamic market in which psychological associations try to regulate test use and agree on norms among psychologists (DIN, 2002; Evers, 1996; Kersting, 2008a), while at the same time non-psychologists look to improve professionalization in the HRD field (Carliner & Hamlin, 2015; Chalofsky, 2007; Lee, 2001; Short, 2006). As different stakeholders have different interests, tensions exist between HRD professionals and professional associations, as well as between psychologists and non-psychologists using personality tests in HRD. Previous exploratory research indicated that professionals place more value on perceived ease-of-use than on psychometrics (Lundgren, Kroon, & Poell, 2017). Also, professionals and publishers re-label the word “test” with more positive-sounding terms, such as “tools for self-reflection and instruments for personal stocktaking” (p. 215) when psychological tests are used for development rather than selection. This indicates that some professionals may have developed strategies that prevent their entanglement with personality tests from being negatively perceived by test takers and client organizations. HRD professionals’ meaning-making (Mezirow, 1991; Schön, 1983) and organizational sensemaking (Greeno, 1997; Maitlis & Christianson, 2014; Weick, 1995) hence seem to play a crucial role in the personality testing industry as they introduce personality tests into organizations or connect test publishers and their products with test takers. The question arises how HR practitioners in general and HRD professionals in specific react to and create meaning from their personality test use.

A number of studies have analyzed HR practitioner reactions to psychological testing in general. In an empirical study conducted with HR practitioners, Benit and Soellner (2013) wanted to find out why personality and intelligence tests are less widely adopted in Germany, compared with other European countries. The researchers collected survey data from 116 companies and found that nearly 20% of those practitioners used personality tests in developmental contexts, mostly in leadership development. Two interesting findings emerged: Practitioners rejected personality tests because of the low face validity of certain tests; however, when they did decide to purchase tests, they would do so from external vendors rather than from psychological test publishers in 43% of all cases. A similar survey study had been conducted in the UK, where practitioners were asked about their reactions to work-related psychological tests (Furnham & Jackson, 2011). In this study, 255 HR practitioners responded to 64 statements that covered their reactions to psychometric tests in general, including cognitive ability (intelligence), aptitude and personality tests. Different from the German sample (Benit & Soellner, 2013), HR practitioners from the UK gave a positive evaluation of personality tests' validity and their usefulness in HR practice (Furnham & Jackson, 2011). Furnham and Jackson (2011) also found that age and educational qualifications of test practitioners were positively related to perceived usefulness of psychological tests. Although the study was conducted in the area of personnel selection and not personnel development, it seems logical that cognitive skills associated with age and educational qualifications are closely linked to understanding the more abstract nature of psychological tests. Furnham and Jackson (2011) further concluded that younger professionals or those with fewer years of education might have a more limited understanding of psychological tests.

While these two survey studies give us a first impression of different HR practitioner reactions to test use, they do not detail the subjective experiences and meaning-making processes that might have led to viewing the practice negatively or positively. We therefore reviewed two more studies that illustrated personal accounts of HRD practitioners and their relationship with personality tests. Ford and Harding (2007) published a reflective analysis of a leadership development program in which the MBTI was used as a tool. The authors take a critical management stance when they refer to MBTI feedback as “similar to horoscopes” (Ford & Harding, 2007, p. 483) and engage in sensemaking based

on personality test use as “means of controlling...individual identities with the organization” (p. 484). In comparison, Scharlau (2004) published a review of her own HRD practice wherein she ascribes meaning to the MBTI as a “useful tool for coaching and career counseling” (p. 13). Through Scharlau’s (2004) frame of reference, the personality test supports the dialogue between coach and client that allows self-esteem to grow and that can be used as a common base for individual development.

In summary, studies from different countries have shown a range of reasons for test use (Furnham & Jackson, 2011) and nonuse (Benit & Soellner, 2013). Furthermore, we understand that test reactions depend largely on the individual and his or her organizational context – some who are critically reflective (Ford & Harding, 2007) and others who are happily embracing the same personality test in their HRD practice (Scharlau, 2004). These opposing reflections on personality test use by HRD practitioners suggest that individual as well as organizational influences determine whether and to what extent personality tests are used in developmental contexts. What is lacking is a theoretical foundation in individual meaning-making (Mezirow, 1991; Schön, 1983) and organizational sensemaking theories (Greeno, 1997; Maitlis & Christianson, 2014; Weick, 1995) that could help us explain the widespread and sustained use of personality tests in developmental contexts among HRD professionals. In this study, we therefore seek to understand how HRD professionals construct meaning and engage in sensemaking, especially where test criticism might lead to conflicting cognitions within the professionals’ practice and among stakeholders of their organizations. By letting participants of this study recount their experiences and their reflections on those experiences, we intend to distinguish among different approaches on how to introduce, administer, and reason the use of personality tests in their practice.

Theoretical foundation

Two key notions inform and shape the foundation of this study: 1) the assumption from the constructivist paradigm that HRD professionals engage in *individual meaning-making* and 2) the idea from the situative paradigm that these professionals participate in *organizational sensemaking*. Both notions will be further elaborated upon below.

Individual meaning-making

To explore the reasoning of HRD professionals, we first turn to meaning-making theories that put the individual at the center of constructing and framing their own experiences. Individuals bring their life and work meanings as well as their knowledge, skills, beliefs, motivation and interests when they reflect on their work practice (Marsick, Watkins, & O’Connor, 2011). They do so by using frames or “meaning perspectives” (Mezirow, 1991, p. 61) that are influenced by social norms, cultural and language codes, prototypes, and philosophies. These sociolinguistic meaning perspectives may represent the individuals’ tapestry of habitual expectations, and they can be seen as the individuals’ frame or paradigm from and through which they observe work practice.

Schön (1983) describes meaning-making frames that provide a set of action-governing operating assumptions as *theories-in-action* – what actually governs people’s actions – as opposed to what people think or say governs their actions (*espoused theories*). Here, the individual constructs meaning from examples, models and metaphors and attaches specific language and descriptors when articulating his or her lived experiences. In *The Reflective Practitioner*, Schön (1983) uses framing terminology to depict how an individual reviews, dissects and reconstructs professional knowledge to take on the character of a system: “The problem he sets, the strategies he employs, the facts he treats as relevant, and his interpersonal theories of action are bound up in his way of framing his role” (1983, p. 210). This line of thought can be applied to personality test use where HRD professionals review, dissect, and reconstruct elements of one or multiple personality tests to make them fit into their workshop practice in management development.

In line with tenets of constructivism (Piaget, 1950; Von Glasersfeld, 1984), individuals build their frames based on subjective knowledge and experiences, from which they extract interpretations. While attributing meaning, (organizational) cultures play a role in the socially constructed realities of these individuals; however, meaning construction is a cognitive process that happens in the individuals’ heads rather than primarily in interaction with the social context.

The constructivist perspective of individual meaning-making is challenged by scholars who argue that social interaction in groups and in organizational

debates shapes individual reflective thinking (Raelin, 2002; Vince, 2002) and should therefore not be neglected. As Clark (1995) elaborates on practitioner engagement in the process of problem solving in the specific setting of teacher education, “the conversation between the practitioner and the setting provides the data which may then lead to new meaning, further reframing, and plans for further action” (1995, p. 245). Dialogue and interaction with colleagues and tools can broaden and deepen the understanding of professional practice. As Yanow & Tsoukas (2009) highlight when reframing Schön’s argument in phenomenological terms, “both self-understanding and evaluative components are learned through engaging in and with the practice, not through thinking about them” (2009, p. 1344). These self-understandings are not qualitatively neutral as they mirror normative concepts used in professional practice and hence also carry with them an evaluative component of right or wrong.

Therefore, the embeddedness of mind in social practice – and the construction and deconstruction of frames in interaction with people, structures, and tools (and hence an extension of meaning-making into the organizational realm) – can make a useful addition that we will consider next.

Organizational sensemaking

Various definitions of sensemaking exist that place the activity of working to understand issues or events that are uncertain or ambiguous as occurring within or between individuals (Weick, 1995; Maitlis & Christianson, 2014; Weick, Sutcliffe, & Obstfeld, 2005). Sensemaking is a social process that occurs among people who negotiate, contest and mutually construct meaning. Sensemaking is also embedded in the professional environment, with the intent to organize and having an ambition for decision-making outcomes. In the context of our study, we understand “organizational sensemaking” (Maitlis & Christianson, 2014, p. 66) as a primarily collective process that happens among individuals and within organizations. As Weick et al. (2005) define it:

Sensemaking unfolds as a sequence in which people concerned with identity in the social context of other actors engage [in] ongoing circumstances from which they extract cues and make plausible sense retrospectively, while enacting more or less order into those ongoing circumstances (p. 409).

We hence turn our focus on the interaction of organizational members and are curious about the way that they interpret their environment and construct accounts that allow them to comprehend the world around them. Here, sense-making can be strategic when it points toward the implementation of organizational change. In a study of middle managers, Rouleau (2005) looks at the processes of *sensemaking* and *sensegiving* through the application of tacit knowledge and finds that these middle managers apply a number of “micro-practices” that help communicate and justify the change. Rouleau concludes that these sense-making micro-practices are socially constructed through nonverbal, implicit meaning structures of activities and words. Organizational sensemaking can therefore be seen as situated in the interaction of organizational stakeholders’ practices, also referred to as the “situative” perspective of learning (Brown, Collins, & Duguid, 1989; Greeno, 1997; Lave & Wenger, 1991).

Viewed from the situative perspective, individuals participate in and collaborate with work-based tools and systems through interaction with their context. According to Greeno (1997), thinking and sensemaking are important aspects of social practice, “involving reflection and discourse on activities of individuals and groups and of meanings of concepts that are significant in evaluating and making sense of the community’s and of individuals’ activities and experiences” (1997, p. 97).

In a study involving Swedish police officers, Lundin & Nuldén (2007) explore how professional tools can trigger workplace learning and how police officers talk about (police) tools and their use. The researchers chose a qualitative ethnographic approach that involved observations and field interviews. Their findings show how the application of specific tools leads to conversations among police officers and how these conversations form an essential part of community-based learning by police professionals. Sensemaking and learning take place in the participation of and interaction between individuals and their environment – a context of rules, values, colleagues, and tools. The researchers conclude that “practice is remembered through the use of tools” (Lundin & Nuldén, 2007, p. 222).

In another situative perspective study that investigates how an anesthesiology department uses organizational practices to help novice nurses become reflec-

tive practitioners, Jordan (2010) inquires about the sociocultural and political context of the individuals' practice. Rather than studying the intrapersonal processes of cognition and meaning-making, the researcher focuses on the processes of participation and interaction that are involved in becoming a (competent) nurse. Adopting an ethnographic approach that includes observation via job shadowing, narrative interviews and document analysis of standard operating procedures, the study concludes that "reflective practices are first and foremost social practices, that is, not certain isolated techniques that are individually applied, but rather interactive ways of approaching and handling situations embedded in a specific organizational and social context" (Jordan, 2010, p. 409).

Fenwick (2016) posits the term "sociomaterialism" to describe a more systems-oriented approach of sensemaking. In her view, professional responsibility is interrelated with materials in motion, for example, in the form of "technologies and texts, objects and bodies, built settings and natural forces" (Fenwick, 2016, p. 167). So whereas meaning-making is portrayed through the constructivist lens as intrapersonal and hence describes the process that happens within oneself when reflecting *in* and *on* action, sensemaking adds a dimension of interrelatedness with people, organizational structures and tools of practice that are situated in the professionals' environment. In the personality testing context, there is a material component, that is, the test itself, as well as a social dynamic component that happens in the interactions of HRD professionals with various other stakeholders before, during, and after delivering a management development workshop. By studying the "inter-acting dynamics of person, social and material elements" (Fenwick, 2016, p. 168), these organizational elements help us understand individuals' entanglement with their social environment and its material elements, for example, the tools that she or he uses for professional practice.

Application of the theoretical foundation to this study

Based on this theoretical foundation, the aim of the present study is to explore how HRD professionals use personality tests in developmental settings. Specifically, we want to find out:

How do individual meaning-making and organizational sensemaking theories help to explain the widespread and sustained use of personality tests among HRD professionals?

We initiated the primary research inquiry to explore how HRD professionals engage with personality tests as tools of their professional practice. We seek to understand how they construct meaning and engage in sensemaking, especially where test criticism might lead to conflicting cognitions within their practice and across stakeholders of their organizations. By letting participants of this study recount their experiences and their reflections on those experiences, we hope to distinguish among different strategies how HRD professionals introduce, administer, and reason the use of personality tests in their practice.

This inquiry is relevant as HRD is a dynamic and constantly evolving field (Lee, 2001) in which the drive for professionalization is apparent (Carliner & Hamlin, 2015) but not well defined (Kahnweiler, 2009). It could be that professionals who continue using personality tests in HRD have adopted strategies for how to deal with test criticism and stakeholder concerns – an area that is worth exploring further, possibly to reduce the research-practice gap in this field (Benit & Soellner, 2013; Short, 2006).

Methods

In order to probe HRD professionals' cognitions on the use and criticism of personality tests for employee development, we followed a grounded theory approach (Glaser & Strauss, 1967) involving semi-structured interviews and inductive data analyses as suggested by Miles and Huberman (1994). We chose this approach in order to develop an explanation of the phenomenon – not merely a description of it. Rather than starting from preconceived theoretical ideas, we engaged in conceptual sensemaking while immersed in the data (Glaser, 1998). Grounded theory guided us in this process of discovery of theory that was inherent in the data we collected, which we analyzed going back and forth using constant comparison together with participants in this qualitative study.

Sample

By means of a purposive sampling strategy, we searched for professionals who were experienced with administering personality tests in HRD, who worked either internal to an organization or delivered consulting/coaching services, and who use personality tests in developmental settings, for example, in management training, team development or executive coaching.

Our process of discovery had started with an interview in 2012 in which we noticed a link between an HRD professional's past experience and her own practice of administering personality tests in developmental contexts. After listening to this test taker's story – we gave her the pseudonym Katie – we started looking for more HRD professionals who were willing to talk about their experiences, good or bad.

We made use of personal and professional networks to support the collection of data. An initial group of HRD professionals who were known to the authors were approached first, thus facilitating cognitive access and creating a level of trust with research participants (Anderson, 2017). Our sampling then radiated outward by asking each interview partner to recommend a colleague or collaborator to participate in the research. Additionally, a request for research participation was posted on selected LinkedIn groups, including “HR Users of Psychometrics” and “The Psychometrics Forum,” which yielded little success. Individuals working for test publishers such as OPP or Hogan or for professional association such as the British Psychological Society were excluded from participation as we thought their commercial interests might bias our research findings. The aim of our purposive sampling was to maximize variation of characteristics among study participants in terms of gender, age, professional experience, exposure to different personality tests, academic background, and position in the company.

As part of a previous study (Lundgren, Kroon, et al., 2017), we had investigated the social contexts of HRD professionals and the personality testing industry through observation, focus group discussions, semi-structured interviews and the collection of personality test feedback reports. Hence, we felt that we had familiarized ourselves sufficiently with the context and setting, which – after conducting 18 interviews and themes and approaches started repeating them-

selves more and more often – led us to notice that we had reached a satisfactory level of saturation (Anderson, 2017). The data collection for this paper was concluded in late 2016 by a team of researchers of whom one is based in the US and the other two are based in the Netherlands. The US-based researcher grew up in Germany and had previously worked in the UK for a number of years, allowing for a somewhat international yet Western-oriented perspective on personality test use.

The sample consisted of 18 HRD professionals in different roles and from diverse industry backgrounds, with slightly more male (56%) than female (44%) participants. Ages varied, with half of the participants falling into the age group 45-54 years. The majority of respondents held a master’s degree (61%), many of whom had a background in psychology (44%), business (17%) or organizational development (11%). Only one participant reported that she never used personality tests in her practice (6%), in comparison with many using tests frequently (44%), occasionally (28%) and rarely (22%) (see Table 4.1).

Table 4.1 Overview of HRD participants

Participants	Job title (Years of test use experience)	Frequency of test use	Personality tests used in HRD practice¹³
Katie	Professional Development Adviser (6)	Occasionally	7; 8
Stephanie	Talent Partner (4)	Frequently	9; 3
Martin	HR Manager (15)	Rarely	7; 11; 4
Nick	Senior Consultant (15)	Frequently	1; 2; 5
Dana	Senior Director HR (4)	Rarely	3; 7; 8
Shaun	Global L&D Director (15)	Occasionally	9; 10; 7
Dominic	Senior Expert (5)	Frequently	9; 7
Hans	HR Business Partner (12)	Rarely	9; 7; 2
Emma	Head of Talent Development (8)	Frequently	9; 7; 3
Saskia	HR Director (12)	Occasionally	7; 11; 6
Lila	Director Talent Management (8)	Frequently	3; 2
Bea	Senior Consultant (21)	Frequently	5; 3; 7; others
Sophie	Education Specialist (-)	Never	7; 10 ¹⁴
Nathan	Independent Consultant (20)	Frequently	12; 9; 10; 7; 8; others
Bernhard	Managing Partner (40)	Occasionally	5; 1
Jackie	L&D Manager (8)	Frequently	7; 2; 8
Reeta	Head of HR (11)	Rarely	2; 8
Bob	VP Organization Development (7)	Occasionally	3; 2; 8

Length of experience with personality tests varied, as did the professional HRD qualifications reported. Participants were located in the USA, UK, the Netherlands, Germany, and Denmark, and upon further inquiry, we found that their country of practice was not necessarily their country of citizenship or origin. Since many participants were international and worked within a broad geographic area or globally, we decided not to focus on national culture as a distinguishing element in our analysis.

¹³ Reference Appendix 1 for test names and descriptions.

¹⁴ Experience as participant

Procedure

An interview guide was developed (see Appendix 3), and semi-structured interviews were conducted that ranged from 35-50 minutes. Because of the international nature of HRD practice, we did not want to constrain ourselves to one geographic area and hence decided to conduct the interviews via phone or Skype in order to broaden our geographic reach. All interviews were audio-recorded, transcribed, summarized, and sent back to study participants for member-checking (Lincoln & Guba, 1985). According to Anderson (2017) “verification of data, analytic categories, interpretations, and conclusions with members of groups from whom the data were originally collected has been accepted as an important technique for establishing credibility” (p. 129). In order to ensure consistency in the interview process, the lead author, who is experienced in leadership development using personality tests, conducted all interviews. We acknowledge that the lead author’s positionality will have affected our research findings, and we will elaborate on this further in the paper’s discussion section.

During the interviews, participants were invited to speak openly about their personality testing experience, their doubts, and their dilemmas. Complete anonymity of interviewees was ensured by using pseudonyms and by deleting any personal, location-based or organizational references from the interview transcripts. Participation in the study was voluntary, and exit ethics were guaranteed. No situational ethics dilemmas were encountered, other than the confession of minor copyright infringements where participants had used training materials that they had not purchased from the copyright holder. In each of these cases, the interviewer clarified how a similar situation could be handled suitably going forward.

While we followed the interview guide that inquired about respondents’ lived experiences, their reasoning and their felt challenges, we acknowledge that our approach to sensemaking is subjective and influenced by the constructivist lens that we adopted, in line with grounded data that consist of the participants’ subjective experiences, interpretations and meanings (Maines, 1991).

Data analysis

An inductive analysis approach was followed to code the interview data. After the first set of interviews had been transcribed, the lead author started to code

those interview sections related to the research question using open coding. After this initial round, all codes were pulled together and grouped within and across interviews. For example, the interview question “What are your concerns when using personality tests in HRD?” resulted in answers that were labeled as open codes, such as “addressing risk of pigeonholing.” This initial coding allowed us to compare interviews and to select those codes that came up more often. When a code was mentioned more than three times, sufficient support was assumed for exploring this code further in the next round of selective coding of all interview transcripts to identify recurring themes (Glaser & Strauss, 1967). This resulted in an initial understanding of participants’ individual meaning-making and organizational sensemaking after this second round of coding.

Next, interview summaries were created for each participant that included passages that directly related to the selective codes. By means of constant comparison, these interview summaries helped to reduce the amount of data for further analysis and to reach out to participants one more time to ask for a member-checking. Of the 18 participants, 12 replied to the member-check request. Of those 12, seven participants requested some small changes to their interview summaries, four accepted without changes and one asked for more time to review the summaries. Hence, feedback from the member-check was available for eleven of the 18 participants.

In the third round of coding, theoretical codes were used to seek commonalities and sort out relationships among the HRD professionals’ various approaches to personality test use. Original interview transcripts were revisited several times to check that the strategies we established as theoretical codes were accurately depicting the HRD professionals’ accounts (Rocco, 2003; Miles & Huberman, 1994). By looking at similar and dissimilar aspects, the patterns we found seem to address the question of “what goes with what” (Miles & Huberman, 1994, p. 249). This approach resulted in six strategies that helped to reduce the data into analyzable units and answer the research question.

Table 4.2 Strategies in personality test use

Strategy	Descriptors	Selected quote	Participants
1 Ethical-protective	<ul style="list-style-type: none"> • High ethical standards • Psychologically trained • Motivated to protect the profession 	<p>"This is the field of psychologists, and, yes, other people can do it, but not in my organization" (Lila, 230).</p>	Lila
2 Scientific-selective	<ul style="list-style-type: none"> • Use of more complex/costly tools • Integrate negative aspects • Offering a portfolio of assessments • Motivated to select best-fitting tool 	<p>"I like the Hogan because it is more in depth, and I think helps people to really understand on a deeper level, because people are always multi-faceted" (Dana, 72-74).</p>	Dana; Emma; Nick; Bernhard; Stephanie; Bea
3 Cautious-avoiding	<ul style="list-style-type: none"> • Skeptical toward testing practice • Negative personal experiences • Motivated to avoid pigeonholing 	<p>"We need to be very careful, because people can take the result of the assessment as a negative sign for them, for their career" (Hans, 199-201).</p>	Sophie; Hans
4 Cautious-embracing	<ul style="list-style-type: none"> • Skeptical toward testing practice • Personal experience mostly positive • Motivated by "light use", e.g., as conversation starter 	<p>"A tool is a conversation starter with a value and purpose...you have to decide how to exploit that for a greater good" (Reeta, 232-234)</p>	Bob; Katie; Reeta; Shaun
5 User friendly-pragmatic	<ul style="list-style-type: none"> • Use of simpler tool that is easy for test taker to understand • Works within organizational constraints • Motivated by high test-taker acceptance 	<p>"The color test worked the best, because people find it easiest to talk about colors" (Saskia, 143-144).</p>	Jackie; Saskia; Martin
6 Knowledgeable-accommodating	<ul style="list-style-type: none"> • Extensive knowledge on industry and tools • Follow ethics guidelines generally, breaking them when demanded by client / employer • Motivated to accommodate clients' needs 	<p>"I said, 'You do know I'm not trained in it.' They say, 'Oh yes, but you are trained in Jung, aren't you?' I thought, 'I've read his books, if that counts'" (Nathan, 138-141).</p>	Dominic; Nathan

Strategies in HRD test use

Professionals make use of different strategies when implementing personality tests in developmental contexts. Each strategy summarizes how HRD professionals apply personal meaning-making and organizational sensemaking structures to their approach when introducing a test, when administering it and when dealing with test criticism and concerns. We found six strategies that professionals employ to frame their use of personality tests in developmental contexts (see Table 4.2).

Ethical-protective

The first strategy is characterized by upholding high ethical standards, no matter what. “Doing the right thing” fuels the *ethical-protective* strategy. For example, professionals who adopt this approach are reluctant to share personal profile information with the line manager without prior consent of the test taker. Psychological training might increase a strong sense for protecting the profession when it comes to administering personality tests in developmental contexts. Following this strategy, test accreditation and general psychometric qualification are seen as important enablers.

In our data, we found one HRD practitioner who seems to conduct her practice following the *ethical-protective* strategy. Lila, the director of talent management at a human capital consultancy, feels strongly about how a personality test ought to be introduced and administered in a workplace setting. When asked about concerns, Lila talks about the risk of stereotyping and refers to the Hogan tools which – in her view – do not categorize people: “I think...the risk is higher when you get something like a typology” (Lila, 277-278). She frames the MBTI as one of the most unscientific tests and states that she would never elect to use it (Lila, 280-281). In fact, she compares type tests to the use of “horoscopes,” (Lila, 285) which she finds unethical because of their lack of scientific grounding. Lila is a trained psychologist and stresses the point again and again that only trained psychologists should be allowed to administer psychometric tests. Summarizing her point of view, she states, “This is the field of psychologists, and, yes, other people can do it, but not in my organization” (Lila, 230). In order to substantiate her view on personality test selection and use, Lila draws an analogy: When you get sick, you consult a doctor and not a carpenter for medical help or advice.

Similarly, Lila explains that only psychologists should handle personality tests, as psychologists have the expert knowledge to choose the right assessment for a given setting.

When inquiring about Lila’s source of skepticism, we found out about her personal experiences when she was first introduced to personality testing. After completing the Hogan assessment for the first time, she found herself in denial about her own test results on one of the assessment scales: “The results looked more...like a cruel monster” (Lila, 89). Lila had been quite skeptical toward that tool beforehand, and when she got her test feedback, she grappled with it further. “I put it in my desk, and after two or three weeks looking again at it, I started to think about, okay, yeah, maybe yes” (Lila, 112-114). Because of her own experience, Lila now starts her workshops with an introduction that she believes helps participants to deal with negative test feedback. She explains to them that it “is a normal reaction that you may deny something” (Lila, 76-78) to assure workshop participants and encourage openness for test feedback.

Scientific-selective

While drawing from psychometric test qualities and scientific evidence, the focus of the second strategy is on choosing the most robust tool rather than being selective about who gets to administer it. HRD professionals are often certified in three or more personality tests, and it is not uncommon that they select more complex tools as part of their professionalization. These chosen tests may also be more costly and elaborate, which may represent higher quality and greater levels of sophistication. Equally, negative personality traits like the “de-railers” in the Hogan assessment are perceived to give more depth and are welcome in the *scientific-selective* strategy. This approach can be further characterized by taking pride in offering a portfolio of tools and by selecting the appropriate one depending on the situation and developmental purpose.

In our data, we found six HRD practitioners who appear to lead their practice following the *scientific-selective* strategy. For example, Bernhard, who works as a managing partner at an executive search and leadership consultancy, feels strongly about type tests, like Management Drives, DISC, and other variations of the MBTI personality tests that are based on Jung’s typology. He calls them the “*kleur terreur*, the terror of color” (Bernhard, 131) and adds that he wouldn’t

use the MBTI as it labels people. He explains that it is not useful for developmental reasons to use “color as an alibi for not being developed” (Bernhard, 136-137) – a point made by other professionals who frame their practices as *scientific-selective*. Bernhard believes that the practice of labeling people actually hinders development. However, he also acknowledges that many of the type tests are popular due to their “enormous good marketing” (Bernhard, 141) and because they are “easy to understand” (Bernhard, 143). However, for his own practice, Bernhard prefers to be selective and only chooses instruments that have a strong scientific foundation.

Bea, a senior consultant at the same executive search and leadership consultancy as Bernhard, is generally open to employing a variety of tests in her practice because she uses them “to start a reflection process with people, and to open a dialogue, and to see what does this mean for you and how does that work for you” (Bea, 66-68). Bea would not base a development center decision purely on a personality test result. The test, as she confirms several times, is the “starting point of a discussion” (Bea, 71), and she is accredited to offer various personality tests. Bea explains her preference in test selection by saying, “I like the Big Five. I’m not so much a type questionnaire person” (Bea, 83-84). When working with type tests, Bea feels limited in the discussions she can have, as type tests do not allow her to conduct deep conversations about what the results mean for the test taker and what their behavior looks like in practice. When a client asks Bea to use a type test like MBTI or Insights, Bea tries to accommodate the request as much as she can. “It depends a little bit on what they want me to do with that” (Bea, 179-180). However, Bea also explains that she will often find a “good reason” (Bea, 180) why the company should rather do the intervention using a different tool, one that Bea recommends. That way, Bea can keep her selectivity on tools, and she can also uphold integrity with what tests she approves as scientific.

In a similar way, Dana likes the Hogan as “it is more in depth, and I think helps people to really understand on a deeper level, because people are always multi-faceted” (Dana, 72-74). Dana works as a senior HR director at an industrial equipment company, where she prefer Hogan assessments that portrait people as “multifaceted” (Dana, 74); this is in comparison with other profiling tools that can be “a bit superficial” (Dana, 74-75). In the past, she has used other tools

with teams, for example the TMP, and she remembers that eight of nine people came back as the same type, which did not help to explore difference and agree team development opportunities. “It basically just said, we are all the same; you should get along and work well together” (Dana, 83-84). Dana concludes that the success of an intervention has to do with the quality of the tool that is used. Dana confirms that she finds the depth and accuracy of the Hogan assessments positive. However, she also sees some downsides to the tools in terms of portraying “negatives” or aspects that can be viewed as negative by the test taker. This is quite different from the TMP or MBTI, in which the describing categories are rather broad and always positive.

Emma, the head of talent development at an industrial equipment company, clarifies that in her view, a “person is never equal to the assessment” (Emma, 255). She explains that “human beings are much more diverse and rich and unpredictable and varied and wonderful, and we can never ever capture everything that is you as a person in an assessment” (Emma, 255-257). What the personality test can do is to give the team a “recognized language...in logical terms” (Emma, 258-259) that explains “what makes our personalities and how...we interact as social beings with other human beings” (Emma, 259-260). Emma believes that test criticism is valid and an HR business partner should always make sure there is enough time to give feedback “because otherwise it just becomes one of those things that HR pulls out of the drawer and then we think everything’s fine and it may not be” (Emma, 268-270). This statement shows that Emma reflects about her professional practice and that she can articulate the limits of a personality test when used in development contexts. Emma adds that “learning is not necessarily going to be nice and linear, and smooth and sexy” (Emma, 383-385) and explains that participants could be learning the most when they are allowed to experience being outside the comfort zone, which can be triggered by personality tests or other developmental tools.

Cautious–avoiding

As a third strategy, we experienced that some professionals are quite cautious about their use of personality assessments, thus avoiding or minimizing the use of personality tests in developmental contexts. Overall skepticism stems from negative personal experiences with some tools. A perceived high risk of pigeonholing leads to cautious behavior around personality tests.

In our data, we identified two HRD practitioners who seem to conduct their practice following the *cautious-avoiding* strategy. Sophie, a learning and development (L&D) adviser who works in public administration, experienced the use of personality tests in a negative way in that she “found the model limiting and misused” (Sophie, line 399). Sophie explains the situation that she was in:

I thought of another method that I experienced in my last job that was of the Insights model, which divides people into four types/colors...I was typed a blue type and therefore my manager really stereotyped me. By this, I mean that he told me several times (in meetings or face-to-face) that I should try to be more of a different color because I would make the working situation more difficult or less pleasant. He was bright yellow and wanted people to try to be as positive and less organized (388-389; 395-399).

This negative experience has influenced Sophie’s professional practice, and she now avoids the use of personality tests in her work in L&D.

Also using the *cautious-avoiding* strategy, Hans understands that personality tests can be used in the wrong way and states that “we need to be very careful, because people can take the result of the assessment [as] a negative sign for them, for their career” (Hans, 199-201). Hans, who works as an HR business partner at an industrial packaging company, refers back to an incident he witnessed in which a manager looked at the feedback report during a development discussion and said, “Look, I don’t understand how you are able to work with all of that. We need to fix it urgently” (Hans, 198-199). Hans was shocked by this statement and decided to adjust his professional practice accordingly. He now prefers to take a softer approach when looking at areas of development based on test results; he believes that employees otherwise develop a “very defensive attitude” (Hans, 207) which then makes the conversation “just offense and defense” (Hans, 208).

When asked about personality test choice, Hans talks about the DISC test and how he finds it unsuitable for developmental purposes: “In general, a DISC test for me is not a reliable source of information because it changes over time and it depends on the mood the candidate was feeling, during filling the questions” (Hans, 38-40). Hans’s negative experiences seem to have influenced his profes-

sional practice, and Hans approaches this subject area with much more caution than other HRD professionals who were interviewed in this study.

Cautious-embracing

A variation of the previous approach, the fourth strategy – named *cautious-embracing* – describes a “light-use” adoption of personality tests in the workplace. Different from the *cautious-avoiding* strategy, having more positive experiences with personality tests – or a turnaround from negative to positive experiences – fuels this approach. The personality test is not seen as a test or assessment in this approach; it is rather a conversation starter and a learning instrument that fosters self-awareness.

In our data, we identified four HRD practitioners who appear to lead their practices based on the *cautious-embracing* strategy. Katie, a professional development adviser in higher education, has experienced the use of personality tests in a similar negative way as Sophie and Hans, as described in the previous section. At some point in her career, Katie was part of a team-coaching session with MBTI that was handled badly in Katie’s view. The external consultant took sides by revealing her own profile, therewith polarizing the group even further. “The outcome was that the others were completely the opposite of me or if there was one letter difference...They were highly introverted, whereas I was highly extroverted” (Katie, 109-112). The team coach positioned herself with all the other colleagues on the one end of the spectrum, and Katie felt stigmatized because her personality test result was different from the majority of the group. “I was just very upset. I said, ‘I can see perfectly well why I don’t fit in here’ but the consultant I think she never should have revealed that she was also ISTP” (Katie, 143-144). Later on, Katie became a management development coach herself and has since tried to support “preference minorities” better in her own workshops. Katie’s own experience has shaped the way she introduces and uses personality tests in her professional practice; she uses an approach that emphasizes test results less and works more with the dynamics of the team. About the use of personality tests, she summarizes, “I do it light” (Katie, 409). Katie’s account explains how her negative experience has led her to become extremely cautious about personality test use and how she has managed to reframe her own practice by following a light-use approach.

Bob, the vice president of organization development at a technology services company, equally expresses caution in situations when people “sing about something” (Bob, 265) that matches exactly what they wanted to hear rather than embracing something that helps them change and adapt. Bob acknowledges that some people love psychometrics so much that it almost becomes “a gospel” (Bob, 245) to them to the extent that they start being “evangelical” (Bob, 245) about their test feedback. Bob has worked with groups who are more pragmatic and who actually feel strongly against the people who do get evangelical about it: “If you get a leader running around saying, ‘I’m ENTJ,’ you look at him and go, ‘Stop talking crap.’ Talk to me like you’re a real person” (Bob, 273-275). Bob feels that he himself falls into that category of people who cannot easily tolerate those who are evangelical about personality tests, and he therefore follows the *cautious-embracing* strategy in his own practice.

Reeta, who works as head of HR at an industrial wood products company, points out the importance of selecting the right person to deliver the personality test feedback. She explains that managers can sometimes take the test result as “absolute gospel” (Reeta, 138) and that the tendency has two sides: On the one hand, it is great to see that people can identify with the test result. On the other hand, managers might “start performing perhaps even more in an extreme version of themselves” (Reeta, 140-141), a tendency that Reeta sees as critical and requiring of a skilled facilitator. Reeta summarizes, “critical reflection is good but again is down to the skill of the person giving feedback and how they deliver this to ensure the best possible environment” (Reeta, 161-162).

Reeta brings up another interesting aspect when she warns professionals that they “really have to watch for people pigeonholing themselves” (Reeta, 57-58). What she means by self-pigeonholing is the situation when test takers take MBTI or TMP feedback from the personality test too literally. Instead, Reeta encourages people to look beyond the immediate test outcome, and she also stresses the fact that personality tests measure preferences, not actual behavior. Because of its openness of scales, she likes the OPQ, as it doesn’t “really come out as one ‘right or wrong’ type” (Reeta, 60-61). Reeta believes that a tool is a “conversation starter with a value and purpose,” (232-233) which defines how she uses the personality test in a *cautious-embracing* way.

User friendly–pragmatic

Professionals who pursue the fifth strategy frame personality tests as simple tools that should be easy to understand, are user-friendly, and have high test-taker acceptance. Test feedback that makes it easy for participants to recognize themselves is valued more highly than overly scientific or complex tools. Scientific evidence or test qualities are rarely mentioned as decision-making criteria in test selection. The orientation in this approach is practical rather than theoretical. Its decision-making logic is deeply pragmatic, focusing on choosing tools that are user-friendly and work best from the test-taker perspective.

In our data, we found three HRD practitioners who seem to conduct their practices following the *user friendly-pragmatic* strategy. Jackie, an L&D manager at a research center, encountered the MBTI as her first personality test. She remembers that she connected with the tool quite well because “I was really looking at myself, probably for the first time ever, really honestly. What I got back reflected me” (Jackie, 45-46). Jackie has continued using the MBTI, and she has added a couple of other tools to her portfolio, including the TMP and OPQ. When dealing with criticism from test takers, Jackie explains that the decisive factor is how well a profile describes a person. “If they think it’s accurate, then they don’t mind too much being put in a box. If they don’t think it’s accurate, then often that’s when they resist it” (Jackie, 383-385). With the MBTI, for example, Jackie appreciates that test takers can work out what is a better fit for them in comparison with the self-reported type they receive after completing the questionnaire. She explains that this is not the case for the Insights or TMP assessments, in which the test taker receives a “lovely, printed report, which is nice if that reflects how you see yourself; but actually, if you get something else back, then that’s when, I think, the resistance often sets in” (Jackie, 388-390).

Saskia, the HR director at a consumer good company, follows the same pragmatic approach. Before moving to the corporate office, she worked as an HR business partner at a manufacturing site where she used to administer different tests, such as My Motivation, MBTI, and Management Drives. When comparing different personality tests, Saskia finds that My Motivation is accessible for everyone while MBTI is “too sophisticated,” which in her view makes it harder for participants from a production environment to grasp it. Participants find the color test the easiest as they can implement that concept into their daily language:

“Look at your Blue today” or “We need more Red” (Saskia, 145). Saskia finds that the colors help depersonalize the feedback, which makes it easier to give that feedback. Saskia reflects on different tools and why she uses them: The MBTI, despite it being the more research-supported and “clearly more valid test” (172), does not always work effectively, for example, in factory environments. That’s why her team decided to use My Motivation with the focus on encouraging dialogue among people, giving and receiving feedback, and reflecting on own behavior. Saskia confirms that “the color test worked the best, because people find it easiest to talk about colors” (Saskia, 143-144), a statement that confirms the *user friendly-pragmatic* approach that Saskia is taking.

Knowledgeable–accommodating

Finally, comprehensive knowledge of the testing industry characterizes the sixth strategy, including knowledge of products, professional organizations, and standards. Awareness of ethical guidelines exists, and this approach allows for some flexibility when playing “according to the rules.” For example, HRD practitioners pursuing this strategy find it permissible to break testing guidelines occasionally when required in a given situation or demanded by a client. Their reasoning is that a certain degree of flexibility is needed when it comes to accommodating client needs – a focus that shows strongly in this strategy. For example, a professional who generally does not use the MBTI might reason that it is okay to accommodate a client’s need to use the tool despite the professional not being accredited for it.

In our data, we identified two HRD practitioners who appear to lead their practices following the *knowledgeable–accommodating* strategy. Nathan, an independent consultant who has been working in L&D across different industries for just under 20 years, is an insider within the personality testing industry as well as being knowledgeable about its numerous tools. Nathan warns how easily a tool could be overrated and what that could lead to, giving an example of the airline company he used to work for: “They want to do it with lots of teams because they see it as the panacea for everything. You have to be really careful. You have to stick to your disciplines and why you want it...If you’re not careful, people will treat it a bit like [astrological] star signs or something” (Nathan, 24-28). In this way, Nathan is knowledgeable about personality tests and their limits. When selecting tools for a specific client project, Nathan first asks the client to identify

the goal. He then checks what has been done so far. He explains that many companies want to do refreshers of personality profiles that they had done in the past or they want to explore certain development areas further. For example, they might want to look at how a person "comes across" (Nathan, 172) and what his or her "blind spots" (Nathan, 172) are. Nathan's task is to find the right tool that fits with the intervention the client is looking for, because "that makes the conversations easier" (Nathan, 174). Although Nathan is not accredited, he has also run some Myers-Briggs sessions because the clients had asked him to. Nathan did not feel comfortable doing them at first. "I said, 'You do know I'm not trained in it.' They say, 'Oh yes, but you are trained in Jung, aren't you?' I thought, 'I've read his books, if that counts'" (Nathan, 138-141). Nathan explains that he can facilitate his way through anything "if you have to" (Nathan, 142), showing that he accommodates clients' wishes even if he is not accredited in that specific personality test – an example of the *knowledgeable-accommodating* strategy.

Strategy clusters

Looking at the six strategies, we observe a divide between two strategy clusters:

- Cluster 1-2-3: Ethical-protective, scientific-selective, and cautious-avoiding
- Cluster 4-5-6: Cautious-embracing, user friendly-pragmatic, and knowledgeable-accommodating

When comparing years of experience across the two strategy clusters, we find that HRD practitioners in Cluster 1-2-3 have on average two and a half more years of test experience in comparison with practitioners who employ a strategy from Cluster 4-5-6. We did not find a discernible difference between (internal) HR practitioners and (external) consultants as to how their work environment had influenced the choice of strategy cluster. Within Cluster 4-5-6, we notice that all internal HR practitioners follow the 4. *Cautious-embracing* and 5. *User friendly-pragmatic* strategies, in comparison with consultants who could be found to follow more the 6. *Knowledgeable-accommodating* approach.

Discussion

The aim of this paper was to explore how individual meaning-making and organizational sensemaking theories help to explain the widespread and sustained use of personality tests in developmental contexts among HRD professionals. Interviews with 18 HRD professionals were conducted, from which individual meaning-making and organizational sensemaking structures were distilled based on grounded theory and inductive analysis. Through pattern analysis, six strategies were established that describe professionals' practical approaches in personality test use. These strategies illustrate professionals' approaches when introducing a test, when administering it, and when dealing with test criticism and concerns, thus depicting the professionals' action theories. We find that HRD professionals form their "strategy as pattern" model (Mintzberg, Lampel, Quinn, & Ghoshal, 2002), where an approach to test use is realized and recognized over time rather than being planned or intended. Our findings align with the "strategy as practice" concept that describes a situated and socially negotiated activity where practitioners shape their strategy "through *who* they are, *how* they act and *what* practices they draw upon in that action" (Jarzabkowski, Balogun, & Seidl, 2007, p. 6).

We note that the Cluster 1-2-3 describes professional practice that holds a critical stance towards test use. For example, type-based tests are more frequently challenged, and reasons for using alternative trait-based tests are clearly presented. It seems that all three strategies in Cluster 1-2-3 are highly reflective and rooted in cognition and beliefs, similar to the qualities of *double-loop learning* (Argyris & Schön, 1974, 1996) whereby professionals are willing to change mental models in order to adapt their decision-making rules. Practitioners who practice one of these three strategies have on average more years of experience in the field and an educational background in psychology. Also, this cluster of strategies relates to deeper personal experiences and more complex meaning-making structures in personality testing, including the reflection on critical incidents with test takers. The focus in these strategies leans towards "doing the right thing" – a principle strongly anchored in all of these three strategies – with a higher use of trait tests in their practice. These findings are in line with Furnham and Jackson's (2011) survey results which had revealed evidence that age and

educational qualifications positively influenced the perceived usefulness of psychological tests in general.

In comparison, the remaining three strategies in Cluster 4-5-6 tend to be less critically reflective on test use and professional practice in general. The individuals' reasoning and construction of meaning is more similar to *single-loop learning* (Argyris & Schön, 1974, 1996) in which professionals take certain test limitations for granted without questioning, through individual or collective inquiry, their work tools or their use in HRD (Fenwick, 2003). These "less reflective" strategies focus on perfecting test practice or "doing things right" – whether this is in relation to spending less money or to creating more value for organizational stakeholders. In line with what might be considered best practice encountered in Cluster 4-5-6, professionals use a socio-linguistic meaning perspectives that depict personality tests as conversation starters and instruments for personal discovery. The use of language here is remarkable; it seems as if the word "test" would provoke unpleasant reactions with test takers and organizational stakeholders, so this word is better avoided: "This is not a test." These strategies stand out as a cluster, as it is more common to actively defend one's own practice by constructing acceptable reasons to continue doing what these professionals have done in the past. For example, reasons were articulated why it is okay to keep applying simplified type-based tests (while knowing there were alternative tools available) or why it is permissible to administer a specific test once in a while without being accredited (while acknowledging that ethical practice is important for HRD professionals). The entanglement between practitioner and tool seems tight in comparison with the lower intensity of interaction that can be observed in the first strategy cluster.

Cognitive dissonance in test use

This entanglement and the professional tools observed in Cluster 4-5-6 seem at first jarring, however they might be explained by Festinger's (1962) theory of cognitive dissonance, which details how individuals are motivated to remove dissonance in situations where two inconsistent beliefs provoke feelings of discomfort. In the case of 6. *Knowledgeable-accommodating* strategy, a professional who is aware of ethics guidelines in personality test use is put into a dilemma when asked by his or her boss to deliver a workshop involving a test for which this professional is not accredited. As a non-credentialed professional,

the individual ought not to execute that task. At the same time, the thought of saying no to the boss creates a feeling of discomfort, also referred to as “cognitive dissonance” (Festinger, 1962, p. 17). This unpleasant tension motivates the professional to reduce the dissonance by adding or subtracting cognitions. The professional could also change his or her behavior to reduce the inconsistency; however, social group membership, especially when working internal to an organization, might inhibit behavioral change due to the power of normative influence (Matz & Wood, 2005). In the case of this professional, he or she agrees to administer the test, supported by the cognition that the boss has endorsed the decision (“It must be okay – he is my boss”), or eased by the thought that it will be a one-off activity (“It can’t be too bad if I only do it only once”). The theory of cognitive dissonance might thus explain why professionals move from a 2. *Scientific-selective* to a 5. *User friendly-pragmatic* strategy when confronted with dissonance that can more easily be resolved by changing a previously held belief. It could be proposed that cognitive dissonance influences the choice of strategies and hence the interrelatedness of professionals with colleagues, organizational structures and tools of practice. In order to check whether this proposition is true, we need to understand how professionals move among different available strategies. Are professionals who embrace personality tests in a highly cautious way (4. *Cautious-embracing*) more likely to move to 3. *Cautious-avoiding* when confronted with dissonance in the form of test criticism? Also, do less experienced and/or less educated professionals generally tend to be more accommodating, pragmatic, and embracing as they might have been when confronted with less challenging cognitions? Our findings indicate a certain tendency that confirms this proposition; however, we also note that some professionals were younger in age (Sophie) or had “only” a bachelor’s degree (Stephanie) and that these individuals still exhibited a critical mind-set and a great concern with current test use while employing 2. *Scientific-selective* (Stephanie) and 3. *Cautious-avoiding* (Sophie) strategies.

Alternative lens: Paradox in organizations

The “paradox in organizations” perspective gives an alternative lens for interpreting organizational sensemaking findings that may seem irrational and counterintuitive. According to Lewis and Smith (2014), “a paradox perspective shifts a fundamental assumption in organizational theory. Traditional theory relies on rational, logical and linear approaches, whereas a paradox perspective emerges

from the surprising, counterintuitive and tense” (p. 143). Our starting point was to find out how HRD professionals construct meaning from and make sense of their personality test-use practice despite the tests’ known shortcomings – an apparent paradox that motivated us to conduct this research. According to our findings, more internal HRD professionals, such as HR business partners, choose the 4. *Cautious-embracing* and 5. *User friendly-pragmatic* strategies, which might suggest that organizational actors narrow their attention to factors under their control and therefore will embrace and implement tests in a more pragmatic way. They administer tests that are more within their understanding; consequently, they collaborate more closely with colleagues, consultancies and test publishers that apply a similar narrow and tool-based focus. In their drive towards cognitive consistency (Festinger, 1962) and influenced by members of their social group (Matz & Wood, 2005), organizational actors also strive for simplicity (Miller, 1993) by focusing more narrowly on a single theme or approach, or in our case, on a single personality test.

In our study, HRD professionals who tend to follow the 4. *Cautious-embracing*, 5. *User friendly-pragmatic* or 6. *Knowledgeable-accommodating* strategies seem to feel that type tests are more under their control and within their understanding in comparison with more elaborate or complex trait tests. As these practitioners cling to their preferred priorities – in close entanglement with the tools they know best – HRD professionals who tend to follow the 1. *Ethical-protective*, 2. *Scientific-selective* or 3. *Cautious-avoiding* strategies illustrate the opposite polar in this field of professional tension as they clearly differentiate themselves from those who are perceived as less ethical, scientific or cautious in their test use. Hence, the paradox perspective gives us a foundation to understand that psychologists versus non-psychologists and trait-based versus type-based tests are “socially constructed polarities that mask the simultaneity of conflicting truths” (Lewis, 2000, p. 761).

So what could be these parallel, conflicting truths in our case of personality test use in developmental contexts? Maybe it’s true that the “terror of color” (Bernhard) coexists with the “ENTJ gospel” (Bob), that “non-linear learning” (Emma) complements “depersonalized feedback” (Saskia) and that ethical guidelines (Lila) accompany flexible approaches (Nathan) in test administration. It would be too easy to say that that one strategic approach of test-use is

right and the other is not, but we need to acknowledge that HRD professionals are challenged by cognitive dissonances and paradoxical situations in their organizational practices on a regular basis.

Research limitations

We acknowledge several research limitations of this study. The relatively small sample size (N=18) allowed us to conduct an exploratory study that future research studies can build on. We also note that our strategies were not confirmed through statistical analysis. Because of the qualitative nature of this paper, the findings on strategies were constructed based on the parameters available “by hand,” not involving factor analysis. Hence, the strategies described here are the result of an exploratory approach that runs the risk of being overlapping, incomplete or leaving out detail.

HRD professionals who were highly engaged with personality tests in their practice were interviewed in this study. As this was one of the sampling criteria, it is not surprising that all participants interacted with personality tests to some degree and that only a small number of professionals displayed *cautious-avoiding* as their strategy due to our systemic sampling preference.

Lastly, the first author’s positionality and insider bias to the personality testing industry has most likely influenced the way interviews were conducted and data were analyzed (Anderson, 2017). In addition, we are aware that the countries of location of the research team in the Western world will have influenced the analytical outcomes and interpretations of the findings. We therefore systematically and consistently scheduled peer-debriefing sessions, during which the other two authors would challenge and critically evaluate the logic applied to the research and analysis process as well as question the member-checking procedure that was followed. Through these measures as well as the peer review process, we hope to have successfully managed the intersubjectivity in conducting this study and in presenting its findings.

Future research

We encourage follow-up studies to stimulate further debate among HRD professionals on how individual meaning-making and organizational sensemaking structures influence their strategic choices when it comes to personality testing

in developmental contexts. For example, one could look into the practitioners’ key drivers or investigate further the implications of personality test use in developmental settings, for instance, from the viewpoint of the test taker.

Another important area of future research stems from the fact that HRD professionals with active personality test practices were interviewed, but those who had never used those tests or who had ceased to use such tests were not included in the study. In a comparative analysis, users and nonusers could be juxtaposed along the question of how meaning-making and sensemaking theories influence the decision of a specific tool to use.

Although the sample comprised participants with a range of (mostly Western) nationalities, further research to examine the issue in other national contexts, for instance, if and how personality tests are applied in Asia, may be appropriate.

Finally, future studies could take the same theoretical foundation and apply it to HRD tool use outside personality tests. For example, a future study could investigate how individual meaning-making and organizational sensemaking influence the use of 360-degree feedback and self-review tools among different HRD professionals.

Practical implications

How can our findings help improve this specific practice of using personality tests in developmental contexts? First, by becoming more aware of their own meaning-making structures and organizational dynamics in test use, HRD professionals can broaden their approaches and make more conscious decisions when it comes to personality test use in developmental contexts. Second, for client organizations that receive management development services from external consultants, our findings give encouragement to probe these consultants in terms of how they reflect on their practice *in* and *on* action, how they deal with uncertainties and limitations of selected tests, and how they manage cognitive dissonances and paradoxes that exist in organizations. Lastly, for test takers and those who encounter personality tests as participants of management development or team coaching sessions, this is a good reminder to remain open, curious and critically reflective about tools and practices that they encounter in managerial life.

Conclusion

In proposing six strategies, we acknowledge an inherent risk of overgeneralizing and hence pigeonholing professionals' approaches into predefined categories. Aside the emergent nature of these approaches, this is not the intent here; laying out strategies is rather an attempt to make visible those tendencies in personality testing that can be seen in HRD practice, with the aim to make sense and construct meaning to explain the widespread and sustained use of personality tests in developmental contexts among HRD professionals. After all, this is not a test – but an exploration of meaning-making and sensemaking structures designed to illustrate different and perhaps paradoxical approaches to personality testing in HRD.

"This is not a test"

Chapter 5

Pigeonholing or learning instrument?

Test takers' reactions to personality testing in management development¹⁵

¹⁵ This chapter has been published as: Lundgren, H., Kroon, B., & Poell, R. F. (2019) Pigeonholing or Learning Instrument? Test takers' reactions to personality testing in management development. *European Journal of Training and Development*.



Abstract

Purpose: While factors that influence test takers' reactions to personality testing in selection contexts have been well researched, little empirical research evidence exists to determine whether these factors also apply to test takers' reactions in the context of management development (MD). The purpose of this study is therefore to explore what explains different test takers' reactions in the context of MD programs.

Design/methodology/approach: A qualitative longitudinal approach with three phases of data collection was employed, resulting in participatory workshop observations and eleven semi-structured interviews with participants from two different contexts. Data was analyzed using Qualitative Comparative Analysis (QCA).

Findings: The findings show that test takers' reactions vary; some are more accepting, others are more neutral or rejecting, where *perceived usefulness*, *clarity of purpose* and *perceived respectfulness* are identified as distinguishing factors. Individuals also differ in terms of their *awareness of assumptions* and their *perceived emotional safety*, two emerging factors that are relevant in the MD context.

Research limitations: Data was collected during the MD workshops and three months after, but no records of immediate test takers' reactions were included, which could be an addition for future research.

Practical implications: The findings of this study suggest that human resource development (HRD) professionals have significant impact on test takers' reactions when it comes to encouraging self-reflection and learning along personality tests.

Originality/value: This paper adds to existing research by offering insights into factors in MD settings where participants are concerned about aspects of fairness, learning and behavioral change.

Introduction

Personality tests are a commonly used tool in management development (MD) programs. However, MD participants react in different ways when they are presented with personality test feedback as part of a MD program. Some learners are overly enthusiastic about the feedback and its development opportunities (Clinebell & Stecher, 2003; Scharlau, 2004); others are more concerned about privacy and skeptical towards its value and utility (Boyle, 1995; Michael, 2003; Pittenger, 2005). This variety in reactions could interfere with the learning outcomes of MD programs.

A personality test is a standardized measure to capture personality dimensions, often using self-report measures (Allen & Hartman, 2008; Furnham, 2017). In MD programs, a coach or facilitator sends out an online link beforehand and discusses personality test feedback with test takers as part of the development workshop. Test takers' reactions comprise any "attitudes, affect, or cognitions" (Ryan & Ployhart, 2000, p. 566) that an individual might have about the personality testing process. Test takers' reactions matter because positive reactions can lead to motivation and learning – two important drivers that contribute to the "transfer of training" (Gegenfurtner et al., 2009). Transfer of training describes the extent to which the learning that results from a training experience is implemented on the job, which in turn leads to significant changes in work performance (Bates et al., 2014; Choi & Roulston, 2015). Without the acquisition of new knowledge, nothing can be transferred from the MD program to the workplace, and without motivation, test takers are not likely to spend the time and energy to transfer new insights into their jobs. Transfer of training is essential in MD as organizations are looking to grow their employees' skills and competencies through these learning initiatives. While organizations invest in their leaders' development as they expect them to continuously improve their performance, insights gained from a workshop that involves a personality test can be used as a source of information to reflect on own and others' behavior, and to adjust to meet strategic business needs.

In order to achieve those MD program goals, HRD professionals need to understand what explains different reactions to personality testing among individuals. According to existing studies in management selection, test takers' reactions

are influenced by a number of factors, including the accuracy of personality descriptions (Layne & Ally, 1980), fairness and procedural justice perceptions (Gilliland, 1993; Gilliland & Steiner, 2012; Hausknecht, Day, & Thomas, 2004; Konradt, Garbers, Böge, Erdogan, & Bauer, 2017) and perceived face and predictive validities (Schmitt & Chan, 1999).

The empirical research on fairness and procedural justice is substantive and multi-faceted (Gilliland, 1994, 1995; Gilliland, Benson III, & Schepers, 1998; Ryan & Ployhart, 2000). For example, Gilliland (1995) conducted a study with applicants who had recently gone through a job search and hiring process. The study participants were asked to rate critical incidents of fair and unfair treatment during the process. The collected critical incidents were then categorized into a distinct number of procedural justice rules that were correlated with different hiring outcomes. This study shows that successful applicants were mostly concerned with consistency of treatment, while unsuccessful applicants found timely feedback and perceived bias as most concerning. Gilliland (1995) further illustrates how fairness can be improved for example by being transparent in selection procedures during the application process. If fairness is important in the selection setting, it is likely that this factor also plays a role in developmental contexts of personality test use.

However, are there additional factors that play a role in MD but not in selection, for example around the domains of learning and behavioral change? And how do these different factors contribute to the motivation to transfer insights from a MD program into the workplace?

When looking at personality test use in MD, goals, focus, context, facilitation, delivery method and tools differ from personality test use in selection. Goals in MD include individual development (Lundgren, Kroon, et al., 2017), as well as the transfer of training (Blume, Ford, Baldwin, & Huang, 2010) where personality tests are used to increase self-awareness, with the expectation that this will lead to more reflective practitioners. An internal or external HRD professional often facilitates such a test in the MD context, where test feedback is delivered through instructor-led discussions using small group exercises and role-play. In MD, both trait- and type-based personality tests are used (Chamorro-Premuzic & Furnham, 2009; J. Hogan & Holland, 2003; McCarthy & Garavan, 1999).

In comparison, the purpose of personality tests use in management selection is to make a hiring decision. The focus here is on whether the candidate has a good job profile fit. As the applicant completes the personality test in the context of a specific job, an (internal) HR professional administers and facilitates the test. Feedback is delivered in a one-on-one setting as part of the interview debriefing, or if in the context of an assessment center (AC) as part of the final AC feedback. In management selection, trait-based personality tests are more often used than type-based personality tests (Lievens et al., 2002).

When comparing the different test use purposes, some influencing factors of test takers' reactions that apply in the selection context may also apply in the MD context, while others are re-labeled or expressed differently when the same test is used for a different purpose. In addition, these factors specific to the MD context could interfere with the transfer of training of the MD outcomes. The present inquiry is driven by the following research question:

Which factors can explain differences among test takers' reactions to personality testing in the context of management development?

Since this is an exploratory study, no specific hypotheses were formulated regarding test takers' reactions. However, one underlying thought was that test takers might be at risk of accepting personality test feedback given to them in a workshop without questioning the feedback. In conjunction with that, test takers might not have access to research based pro's and con's of personality tests and depend completely on the information provided to them by the test facilitator. This could manifest itself in an information asymmetry between the facilitator and test takers. The purpose of this paper therefore is to explore test takers' reactions to personality tests in order to engage in theory development and the bettering of HRD practice. In terms of theoretical contributions, this paper looks at the extent to which confirmed factors from the selection literature also hold true in the context of MD. In terms of practical contributions, this paper aims to translate relevant findings into practical actions that engage HRD professionals in reflective practice when they administer personality tests in MD. This paper is structured into three parts. First, an overview of research knowledge on factors that impact test takers' reactions in MD as compared with other (selection) settings is given. Second, a summary of findings from

the empirical research with test takers in two MD programs follows. Finally, theoretical and practical implications for HRD researchers, leaders and practitioners are discussed.

Theory: Test takers' reactions

A number of different factors influence the way test takers react to personality testing, as established in empirical studies in management selection (Beermann, Kersting, Stegt, & Zimmerhofer, 2013; Fluckinger, 2010; Fluckinger & Snell, 2016; Gilliland, 1993), leadership development (Harland, 2003) and assessment-based feedback literatures (Atwater & Brett, 2005; Atwater, Brett, & Charles, 2007; Ryan, Brutus, Greguras, & Hakel, 2000). These studies refer to the domains of fairness, learning and behavioral change, all of which link to the use of personality testing in developmental contexts. Although many different influencing factors to test takers' reactions are suggested as part of these domains, not all of these are equally important in the MD context where transfer of training (Bates et al., 2014; Blume et al., 2010; Burke & Hutchins, 2007; Choi & Roulston, 2015) and the application of adult learning principles play a role (Chalofsky, 2007; Fenwick, 2016; Garavan, McGuire, & Lee, 2015). Therefore, only those factors that are closely linked to the MD context were selected for the present study (see Table 5.1). The following paragraphs will discuss each of the listed influencing factors, their source and why they could be relevant when describing test takers' reactions in a MD context.

Perceived measurement quality

In a study among 110 participants with work experience, Beermann et al (2013) first administered a work-related personality test and then asked participants to rate their experience using an acceptance survey instrument that had previously been employed to measure reactions to intelligence tests (Kersting, 2008b) and ACs (Kersting, 2010). The study's results show that perceived measurement quality, that is: the extent to which a test taker believes that the test will generate results that will accurately reflect oneself, is highly correlated with personality test acceptance (Beermann et al., 2013). Measurement quality perception here refers to a belief held by the test taker that might or might not match up with the psychometric qualities of the test. Similar to job applicants forming a belief on an assessment tool, it seems plausible that MD participants,

too, would form a belief about the quality of the development tool that they are presented with. In MD, individuals seek tools that can help them improve their practice, for which an accurate representation of preferences and tendencies seems paramount. Therefore, *perceived measurement quality* was included in the present study.

Table 5.1 Hypothesized factors that influence test takers' reactions

Factor	Description	Source
<i>Perceived measurement quality</i>	The extent to which one believes that the test will generate results that will accurately reflect oneself.	Beermann et al. (2013)
<i>Perceived usefulness</i>	Feedback that can be effectively used to guide decisions related to developmental needs and activities.	Harland (2003)
<i>Clarity of purpose</i>	The extent to which the purpose of the test has been clearly communicated.	Fluckinger & Snell (2016)
<i>Perceived respectfulness</i>	A procedure and interaction that is perceived considerate and appreciative to the test takers' feelings.	Harland (2003)
<i>Self-awareness</i>	The extent to which one knows one's preferences.	Ryan et al. (2000)
<i>Perceived control</i>	The extent to which one perceives the answer format as non-limiting to express oneself.	Gilliland (1993); Harland (2003); Beermann et al. (2013)

Perceived usefulness

Utility-type reaction measures have been found to be strong indicators for changes in learning and work performance improvements in the transfer of training literature (Alliger, Tannenbaum, Bennett, Traver, & Shotland, 1997). If personality tests are perceived as useful feedback instruments to guide decisions related to individual developmental needs and activities, then these tests are likely to receive more positive test takers' reactions. In a study with 255 MBA student, Harland (2003) looked at personality tests with different answer formats in three experimental settings. The author included *perceived usefulness*

as one of the influencing factors and found that different answer formats can impact the test's utility perception. Perceiving the test as more useful can lead to greater transfer of training and hence improved work performance. Since work performance improvement is a key focus in personality testing in the context of MD, *perceived usefulness* was added to the study's list of influencing factors.

Clarity of purpose

Personality tests may vary greatly in terms of what instructions are given to test takers before they complete the questionnaire. In a selection context, individuals might be instructed not to attempt to fake the assessment (Fluckinger & Snell, 2016), which could lead test takers to feel threatened or challenged and hence negatively influence the test takers' reactions (Hough & Oswald, 2008). In a developmental setting, individuals may be instructed to relate their answers to a single frame of reference, e.g., to think of specific work examples when completing the questions rather than referring to situations at home. Test takers are often asked to answer the questions in a natural way, without pondering over answers for too long. In a selection setting, the purpose of using personality tests is more often to determine the job profile fit between applicant and open position. This is not necessarily the case in MD where the purpose might be to increase self-understanding (Cacioppe, 1998) or to improve work performance by applying more effective work behaviors (Fluckinger, 2010). Since the *clarity of purpose* ('Why are we doing this?') might alter test takers' reaction to personality testing in MD, the factor was included in this study.

Perceived respectfulness

In Harland's (2003) test-taker study with MBA students, two common response formats were compared: the normative "Likert-type" answer format and the forced-choice "ipsative" answer format. Since individuals are required to choose a statement in the forced-choice format even if they feel none of the offered options describes accurately their preference, the author argued that forced-choice personality tests could be viewed as hurting people's feelings and hence cause negative reactions (Harland, 2003). In a similar way, if individuals in a MD program cannot express their personality test answers freely, that might have a negative effect on test takers' reactions. In addition, the interaction between test taker and facilitator during the feedback session can influence participants'

perceptions of respectfulness, which was the reason for adding this dimension to the present study.

Self-awareness

The extent to which an individual recognizes one's strengths and weaknesses can generally be referred to as self-awareness. In a assessment-based feedback study with 225 individuals who received feedback as part of a MD program, Ryan et al (2000) hypothesized that individuals high in self-awareness would be more receptive to MD feedback. The authors found partial support for that hypothesis. In personality testing, self-awareness may refer more specifically to the degree that a test taker can name his or her preferences along the various personality dimensions. Along the findings of Ryan et al (2000), a person who is more self-aware might also be more comfortable receiving personality test feedback, whether the feedback is positive or negative. Hence, *self-awareness* was added to this study.

Perceived control

Gilliland (1993) was the first one to apply concepts from organizational justice theory to the personnel selection context in order to understand what impacts perceptions of fairness among applicants. Under the procedural justice header, Gilliland describes various factors that impact applicants' reactions, including the "opportunity to perform" construct (Gilliland, 1993, p. 704) that refers to the applicants' perception that the selection process allows them to demonstrate their knowledge, skills and abilities. Others have named this construct "controllability" (Beermann et al., 2013, p. 44) or "perceptions of test-taker control" (Harland, 2003, p. 288). In her study on test format effects, Harland (2003) found that forced-choice answer formats where individuals are not given an option to reflect a "nonresponse" type choice (such as "not sure" or "does not apply") can negatively influence test takers' reactions to such a test. In a similar way, we expect the extent to which test takers perceive the personality test answer format as limiting to express themselves in MD will have an effect on their personality test reaction, and we hence added *perceived control* as factor to the present study.

In summary, we project that perceived measurement quality, perceived usefulness, clarity of purpose, perceived respectfulness, self-awareness and perceived control will all positively affect test takers' reactions to personality tests in MD.

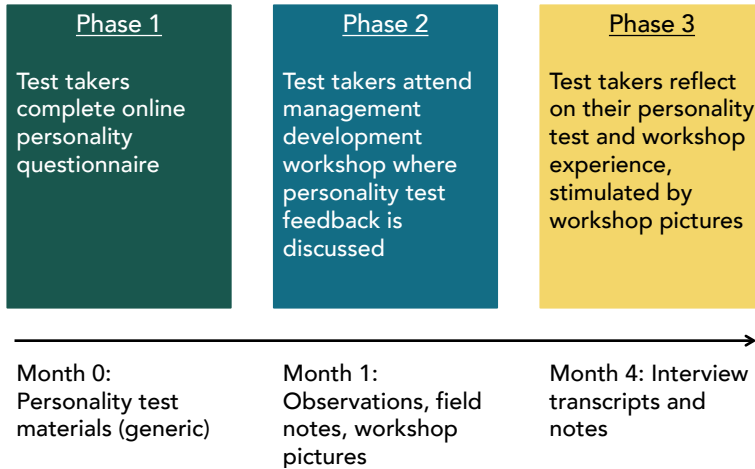


Figure 5.1 Longitudinal design in three phases

Methods

In order to explore differences among test takers' reactions to personality testing in the context of MD, a qualitative approach with a working population, involving participatory observations and semi-structured interviews was employed. This approach was chosen in order to gain insights into the lived experiences of working professionals and their reactions in a natural workshop setting that they attended as part of their MD. Further, Qualitative Comparative Analysis (QCA) was selected in order to conduct a systematic cross-case comparison using pre-determined influencing factors as discriminating conditions.

Design

A qualitative, longitudinal study design with working professionals who were in employment during the study's three phases was conducted (see Figure 5.1).

During phase 1, test takers completed an online personality questionnaire that had been provided by their employer. This happened during “month 0” of the data collection. In parallel, generic materials about the personality tests, e.g., manuals and test reviews, were collected. During phase 2 of the study, which took place one month later, test takers attended a MD workshop where the personality test feedback was discussed. During this phase, field notes and pictures were collected that would later be used during the interviews to stimulate re-call of the event (Fielding & Thomas, 2001). Three months later, during phase 3, test takers were interviewed about their experiences, from which interview transcripts and additional research notes were collected.

Sample and procedure

The sample consisted of eleven test takers who work in different roles for two large organizations in Western Europe. Organization A is a multinational consumer goods company who offers the Team Management Profile (TMP) personality test to their leaders as part of a series of customer service workshops. Organization B is a research-intensive UK university who offers the Myers-Briggs Type Indicator (MBTI) to their employees as part of a four-module MD program. One condition for sampling was that each research participant had to have gone through one of these management workshops; hence they had experienced the phenomenon at hand. The sample consisted of slightly more men (55%) than women (45%), with about half of the participants falling into the age group 25-34 years, one-third 35-44 years and the remaining group 45-54 years. Participants of this study held different positions within their organizations, and most of them were already in line management positions (see Table 5.2).

Organization A was approached first, and after some access negotiations with the functional director of the department, interviews with four voluntary study participants were concluded in October 2012. In parallel, Organization B had agreed to participate in the study, and interviews were conducted with seven voluntary study participants in December 2012 and January 2013. Since the workshops in Organizations A and B had more than 25 participants each, the hope was to recruit a larger number of employees to participate in the present study and to reach a more satisfactory level of saturation (Anderson, 2017). Two more attempts were made to negotiate access with additional organizations from the automotive industry and the public sector, both of which were unsuc-

cessful. The data collection that includes field notes, personality test samples, interview transcripts and participatory observation notes was concluded at the end of 2013.

Table 5.2 Overview of test taker participants

Pseudonym	M/F	Age	Job title	Organization
Thomas	M	45	Manager	A: Multinational consumer goods company; Team Management Profile (TMP) as part of a customer service workshop.
Peter	M	34	Team Leader	
Veronique	F	34	Information Technician	
Kai	M	48	Manager	
Sam	M	25	Medical Doctor	B: Research-intense university; Myers-Briggs Type Indicator (MBTI) as part of a management development program.
Preeti	F	31	Personnel Officer	
Sarah	F	31	Communication Manager	
Alan	M	42	Buildings Maintenance Officer	
Daniella	F	35	Lab Manager	
Lucie	F	39	Personnel Administrator	
Mike	M	40	Head Chef	

Analysis

Comparison lies at the heart of human reasoning as “thinking without comparison is unthinkable” (G. Swanson, 1971, p. 145). In daily life, we use comparison in order to differentiate one object from another. For example, we know that oranges are not lemons because we have compared the two. Comparison is hence a powerful mental operation that can be translated into a set of methods and techniques when used in social sciences research. Because of that, QCA was employed in the present study to conduct a systematic cross-case comparison using pre-determined factors. These factors, sometimes termed “conditions” in QCA terminology (Rihoux & Ragin, 2009), are derived from literature and differentiate cases from one another. Six influencing factors that were extracted from test-taker reaction and feedback literatures were selected (see Table 5.3). This method of comparing the presence or absence of these factors seemed most suitable for the present study as the aim was to explore which of those factors

apply in MD settings. Moreover, comparative analysis is well suited to answer research questions in which a low number of cases is involved (Caramani, 2008), where causal factors can function autonomously or in combination (Blatter & Haverland, 2012). Therefore, a comparative approach was employed to analyze interview transcripts and corresponding notes of eleven cases along the six factors that influence test takers' reactions. This method of cross-case comparison using QCA is useful for small- and intermediate size research designs where sufficient familiarity and empirical intimacy with each case can still be warranted (Rihoux & Ragin, 2009).

Table 5.3 Units of analysis in QCA

	Description
<i>Cases</i>	The number of cases included interviews with test takers (N=11) and participatory observations (N=2). Individuals and organizations are referred to in this study with their pseudonyms.
<i>Factors</i>	Perceived measurement quality Perceived usefulness Clarity of purpose Perceived respectfulness Self-awareness Perceived control
<i>Coding key</i>	"1" for observation where the factor was present "-1" for observation where the factor was absent "0" to indicate that there was mixed evidence "- " to label an instance where the factor was not mentioned

According to Berg-Schlosser and De Meur (2009) each empirical field of study can be described by the cases analyzed, the characteristics of cases being considered, and the number of times each case is observed (see Table 5.3).

The comparative analysis was structured into three steps:

- In Step 1, transcripts and field notes were coded following the coding key and marking all text and observations that related to the six factors.

- In Step 2, a determination was made whether a specific case would be coded as “1”, “-1”, “0” or “-” based on the number of mentions of those factors. Where a determination was difficult, for example whether a case was more of a “1” or a “0”, interview transcripts were reviewed by the group of researchers and in a few occasions the recording was replayed in order to come to a conclusion.
- In Step 3, after the coding determinations, a cross-case comparison was conducted. In order to do so, a table was created to sort those cases that were most similar together, and to contrast them to those cases at the end of the table that were most different.

In addition, open coding and inductive analysis was applied to find additional conditions that had not been derived from literature, resulting in two emerging factors. In the findings section, the results of this QCA will be presented in reverse order, starting with the cross-case comparison before substantiating the findings with interview quotes and field notes. A separate section is dedicated to the two emerging factors and how they related to test takers’ reactions in MD.

Trustworthiness

By discussing issues of transferability, confirmability, dependability and credibility, scientific rigor can be established through the measure of trustworthiness in qualitative research (Bryman, 2008; Lincoln & Guba, 1985; Mertens, 2014).

Transferability acknowledges the “contextual uniqueness” (Bryman, 2008, p. 378) of the social world being studied while looking at the extent to which the findings can be transferred to other settings, which parallels *external validity* in quantitative research. Thick descriptions of the sample, the procedures and the findings were provided, using ample direct quotes and explaining descriptions, all of which contribute to transferability (Mertens, 2014).

In terms of confirmability, which parallels *objectivity*, the research team held back their own experience and knowledge on the subject studied during the data collection phase in order to reduce the risk of attaching meaning and interpretations to the participants’ experiences *a priori*. Although a set of six factors that were theoretically derived was taken as the starting point, the research team

approached the data as openly as possible during coding and analysis, which was discussed and challenged throughout the process.

A research log of all coding decisions was kept and an auditing approach was adopted to manage the dependability of this study (parallels the measure of *reliability*).

Finally, the research team aimed to reach credibility, a description of *internal validity* that demands evidence to ensure that the study was carried out according to the “canons of good practice” (Bryman, 2008, p. 377) in qualitative research. The team did so by engaging in peer debriefing and by sharing preliminary findings of this study with colleagues and the wider HRD research community at various research seminars and conference presentations (Lundgren, 2013, 2014a, 2014b).

Findings – factors that explain test takers’ reactions

The findings section contains two parts. First, findings from the cross-case comparison on test takers’ reactions will be displayed. Next, additional emerging influencing factors that were not previously discussed in literature will be offered for further exploration.

Six factors, derived from literature and discussed at the outset of this article, were hypothesized to distinguish how test takers react when they are exposed to personality testing as part of a MD program (see Table 5.4).

From the cross-case comparison, two findings can be derived. First, some individuals seem to generally perceive the use of personality tests as more positive (see top of the list, e.g., Veronique and Thomas), while others seem less accepting or generally more negative towards its use (see bottom of the list, e.g., Sam and Mike). This differentiation can be made if comparing the number of “1’s” in the top part of the table with the number of “-1’s” in the lower part of Table 5.4. In order to highlight the difference in perception further, two vignettes were selected, depicting a summary of Veronique’s and Sam’s reflection on their respective test-taker experiences. Second, some factors seem to impact positive personality test reactions more deeply, namely *perceived usefulness*, *clarity of*

purpose and *perceived respectfulness*, in comparison to other factors, namely *self-awareness*, *perceived measurement quality* or *perceived control*.

Table 5.4 Results of cross-case comparison

Pseudonym	Self-Awareness	Perceived Control	Perceived Measurement Quality	Perceived Usefulness	Clarity of Purpose	Perceived Respectfulness
Veronique	1	1	1	0	1	1
Thomas	1	1	1	-	1	-
Sarah	-	-1	1	1	1	1
Peter	-	-1	1	1	1	-
Lucie	1	-1	0	1	1	-
Preeti	-	-1	0	1	1	-
Daniella	1	1	-1	-	-	-
Alan	1	-	1	-1	-1	-
Kai	-	-1	1	-1	-1	-
Sam	1	-	-1	-1	-1	-1
Mike	-1	-1	-	-1	-1	-1

The findings will be presented along the three steps of analysis, but in reverse order: First, two contrasting cases will be illustrated using one vignette each before depicting the factors with most direct influence on test takers' reactions. The findings section will also highlight emerging factors that play a role in the context of MD.

Vignette 1

Veronique, 34 years, works as an Information Technician in a corporate environment. After taking the TMP personality test, Veronique was surprised how detailed the personality feedback was, and how well it matched her own percep-

tion of self (17). During the interview, Veronique explains that she has always had an interest in understanding individual differences (39-40) and personality psychology (267-268), and that the workshop has given her positive energy (76) and additional motivation (362). As part of the workshop, Veronique was surprised to see that she was surrounded by many analytical people in her company (123). It made her wonder whether she was in the right job as her preference came out quite different from many of her colleagues. Veronique states that being on one extreme end can be difficult (58) and that she experienced some of the exercises during the feedback workshop as stressful (64). She regrets that for many of her colleagues the personality test feedback was forgotten quickly after the workshop as there was little follow up by her line manager (342). Overall, Veronique found her preferences accurately described in the personality test feedback: “it reminds me of who I am and how I need to be with other people” (205-206). Veronique recommends others to do the profile if they are willing to change and grow (396).

Vignette 2

In comparison, Sam is a 25-year-old medical doctor who works for a university hospital. Prior to coming to the MD workshop of his employer, he had already completed the MBTI as part of a career development workshop (41). He sees the personality test feedback as a useful starting point (29; 413), but finds the overall concept of personality testing challenging for several reasons. First of all, Sam feels ambiguous about his own MBTI results (65-66) and after trying to discuss situational influences with one of the facilitators he “lost faith in the questioning system” (251). Especially the profile descriptions of “J - judging” and “P - perceiving” do not make sense to Sam and are hence confusing for him (154; 168-170; 197). He further found little value in the role-play exercise that was used by the facilitators to show the difference between “I - introversion” and “E - extraversion” (87; 92). In Sam’s view, personality is much more complex than depicted in an MBTI profile and illustrated through workshop exercises (95; 104-105). Sam believes that behavior depends on situational components (194-195; 221-224) and that personality can change over time (209). During the workshop, Sam perceived the facilitator to show disrespectful behavior towards him as he had a discrepancy in two of the four letter: “one of the facilitators said that there can be one letter discrepancy in your two profiles but anything with two or more there is something wrong with you” (252-254). He experi-

enced the personality test procedure as un-scientific (267) and not extremely useful (451). Sam regrets that no suggestions were given during the workshop on how to use profile in the workplace (400-401). It should be noted that Sam and Veronique were not part of the same workshop setting but experienced a different personality test administered by different facilitators and for different MD programs. However, looking at the table of cross-case comparison, Sarah and Sam were part of the same workshop, and so were Kai and Veronique, and yet their experiences were different on many of the dimensions.

When looking at the six factors, three of them, namely *perceived usefulness*, *clarity of purpose* and *perceived respectfulness*, most clearly divide the table into a top half (more positive reactions) and a bottom half (more negative reactions). Mixed evidence for the other three factors in terms on their impact on test takers' reactions were found.

Perceived usefulness

When personality test feedback is effectively used to guide decisions related to development needs and activities then it can be labeled as "useful" in the present study. As discussed in the theory section, perceived usefulness can lead to more positive test takers' reactions that in turn increases the likelihood of transfer of training. A number of participants talked about the utility of personality test feedback, including Sarah, Peter, Lucie and Preeti. For example, Preeti, a 31-year old Personnel Officer at Organization B, recalls that she had been aware of the MBTI for a longer time, but that she had not understood prior to her MD workshop how to make use of it. As she illuminates, "I think it's this report which actually helped me to understand how I can relate my type more to work so that was helpful actually. So it's only after Session One that I started using it" (Preeti, 428-430). In comparison, Kai, 48 years and Manager at Organization A, did not feel that the TMP personality test gave many new insights for him. He was happy to get some feedback, which he found confirming of what he knew about himself already. Kai comments, "Nobody here changed completely" (Kai, 470-471). Interestingly, Veronique, who overall reacted positively to the personality test feedback, did struggle with the question whether she could use the feedback to guide her own development. Mixed evidence of her wanting to take parts of it forward was found while at the same time she found it difficult to implement those things in the workplace.

Clarity of purpose

In the present study, this factor was defined as the extent to which the purpose of the personality test has been clearly communicated by program facilitators. In the cases of Thomas (to recognize development opportunities), Veronique (to create team motivation), Sarah and Lucie (to reflect on self), Peter (to receive confirmation of strengths and weaknesses) and Preeti (to explore learning styles) the purpose had been realized by the test takers. In comparison, there was little or no clarity for Alan, Kai, Sam and Mike as to why a personality test was used as part of their program: “I suppose at first I kind of thought ‘What does that have to do with line management?’” (Alan, 310). Alan further explains that he did not see the relevance of the tool to the rest of the course, and that its purpose only became apparent to him after conducting the research interview as part of this study. Kai describes how the personality test feedback was soon forgotten after the workshop, and that no one in his team really made an effort to pick up the topic during subsequent team meetings. Also Mike, a 40-year-old head chef who works for Organization B, found it difficult to put into practical use the feedback he received from the personality test and “how that fits in with things” (Mike, 22). Similar to Alan, he found the research interview helpful as it highlighted what could have been discussed during the workshop. Sam noticed that the investment in getting MBTI licenses for each participant as well as a certified instructor to run the workshop must have been quite significant, so it was troubling to him that the purpose had not been clearer: “because it’s something that they have invested money in but they never really told us how to use it as a tool” (Sam, 399-401). Sarah, Lucie, Preeti, Mike, Alan and Sam all formed part of the same workshop, delivered by the same set of facilitators, and yet it is interesting to note that the workshop’s purpose with regards to personality test use was clear for some and not for others.

Perceived respectfulness

This factor describes the procedure and interaction that is perceived as considerate and appreciative to the test takers’ feelings. In this present study, comments that hinted towards any respectful or disrespectful behavior or exchanges during the workshops were analyzed as part of this factor. Two test takers mentioned positive interactions (Veronique and Sarah), seven individuals did not comment on this condition, and two test takers felt upset about disrespectful behavior of workshop facilitators towards them (Sam and Mike).

To give a bit of background, both Sam and Mike had been part of the same MD workshop using MBTI at Organization B, and both Sam and Mike had a discrepancy of one or more letters between their self-assess and their reported type profiles¹⁶. As a consequence, both Sam and Mike spoke to the facilitators in a one-on-one setting, and they both felt undervalued and misunderstood during those conversations. During this part of the feedback debrief, Mike was handed a piece of paper with different type descriptions and he was asked to decide which one he was. When Mike tried to explain that at work he was possibly a different person than at home and that he was hence not sure which one to pick, the facilitator reportedly said: “Oh maybe the job you’re doing is not necessarily suited to your reported type!” (Mike, 446-448). Mike feels frustrated about this comment, as he was not given the chance to explain why he felt ambiguous about the profile fit. In a similar way, Sam feels that his interaction with one of the facilitators was frustrating as he was told that one letter discrepancy between the reported and self-assess type profiles was tolerable, but that two or more meant that there was something pathologically wrong with him (254), and that he felt unsure about what to do with this strong statement. In both instances, the test takers did not feel supported and appreciated, which caused them to perceive the interaction with the facilitators as little respectful.

Findings – emerging influencing factors to test takers’ reactions

In addition to findings from the qualitative comparative analysis, the following emerging factors were found to impact test takers’ reactions in MD settings:

- Becoming aware of assumptions
- Perceived emotional safety

Becoming aware of assumptions

By talking about their lived experiences, test takers become aware of assumptions on the concept of differential psychology in general and the dynamics of learning about preference and difference through personality testing during a

¹⁶ For a more elaborate discussion on self-assess and reported type profiles, please see the “Introduction to Type” MBTI manual (Myers, Kirby, & Myers, 2000).

MD workshop in specific. For example, Sam realizes that the personality test he was presented with simplifies the way individuals' preferences are described:

I am probably more of a "P" than a "J" but then again it goes back to that same problem as with the first example. It's not as clean-cut as that. It's just how I approach a task doesn't make me more of a judgment compared to perception – I don't think it's as simple as that (Sam, 191-194).

This process of becoming aware of assumptions while reflecting on lived experiences can be observed with many of the study's participants when they talked about the simplistic nature of tests (Sam; Preeti; Daniella), the difference between behavior at work and behavior at home (Kai; Mike; Lucie) and the fact that personality can change and develop over time (Veronique; Sam).

Perceived emotional safety

During the interviews, test takers expressed how they felt about various workshop parts including the facilitation, exercises and small group discussions. As alluded to earlier, Sam felt uncomfortable about the facilitator's reaction to his personality profile where he had a discrepancy of two letters between his reported and his self-assess type: "It kind of makes you think 'Oh gosh, there is something wrong with you!' probably a personality disorder or something like that" (Sam, 260-261). After this statement, Sam did not feel safe to express himself in this rather unaccepting environment, and consequently his reaction to the MD workshop was more negative. While Veronique also experienced some of the workshop exercises as emotionally exhausting – "It's quite stressing to get your personality out like this before other persons" (Veronique, 63-64) – she felt more assured and respected by the facilitators and hence evaluated her overall experience as more positive. Interesting, the facilitators in both workshops worked hard to build trust and emotional support – possibly to create a space where test takers felt safe. They did so by giving reassurance "You're in good hands" (Organization B, Fieldnotes II, p. 1); by offering confidentiality "You don't have to tell anyone in the room. Your manager will not be told either" (Organization B, Fieldnotes II, p. 12) and by suggesting the limits and boundaries of personality testing "It's not a religion" (Organization B, Fieldnotes II, p. 15). Depending on the level of perceived emotional safety, the reaction to the personality testing part was more positive (Veronique) or more negative (Sam).

Discussion

The aim of this paper was to explore factors that can explain differences among test takers' reactions to personality testing in the context of MD. Through this exploratory cross-case comparison of eleven participants from two settings, three established and two emerging factors were found that help explain why test takers react differently to the use of personality tests in MD (see Figure 5.2).

This study has shown that some influencing factors from selection studies also apply to MD contexts as they link to the domains of fairness, learning and behavioral change: *perceived usefulness*, *clarity of purpose* and *perceived respectfulness*.

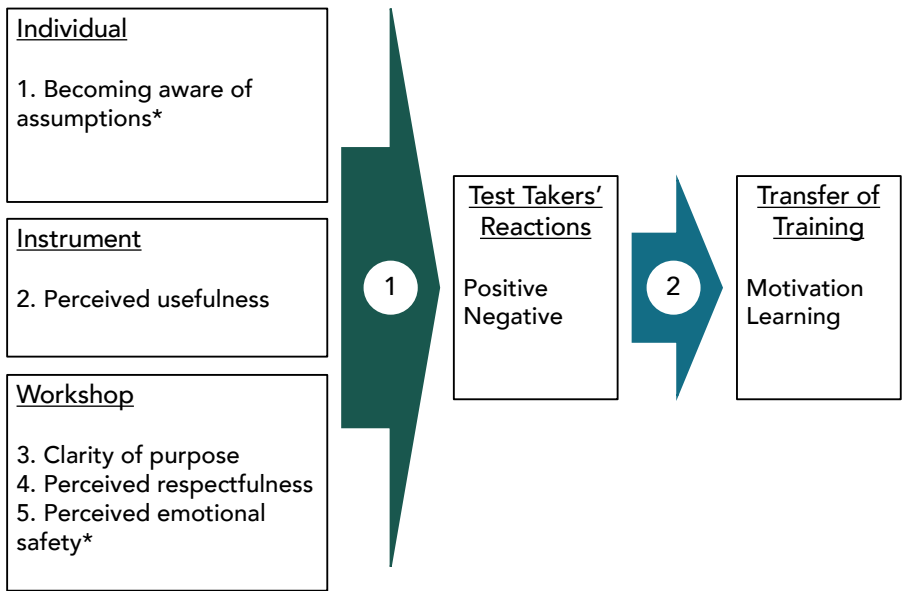


Figure 5.2 Factors that influence test takers' reaction to personality testing

*Emerging factors identified in this study

Perceived respectfulness closely aligns with Gilliland's (1993) procedural justice model, especially the aspect of interpersonal effectiveness that looks at the degree to which participants are treated with warmth and respect. Other factors, such as *perceived measurement quality*, *self-awareness* and *perceived control* play out differently in the MD context and their effect on test takers' reac-

tions could not be established through this exploratory study. Additional factors that may only be relevant for the MD context were found, such as *becoming aware of assumptions* and *perceived emotional safety* where emotions can play a significant role in the learning experience (Raelin, 2001; Short & Yorks, 2002).

Emotions can contribute to or detract from adding meaning to personality feedback received during a workshop, in which the role of the facilitator is to foster learning and development (Gayle, Cortez, & Preiss, 2013; Howie & Bagnall, 2015; Mezirow, 1991). It is possible that participants bring emotions into the training room from the outside (Short & Yorks, 2002), but those emotions generated by the course process as well as through the interactions between and among participants and facilitator can equally affect test takers' reactions. After all, personality testing is a highly personal matter that – when discussed “publicly” in the context of an MD initiative – may cause hope, anger or other emotions that could affect the participants' motivation to transfer (Bates et al., 2014; Gegenfurtner et al., 2009).

When looking at the link between influencing factors and test takers' reactions, the direction is clear for four of these five factors: more leads to more positive reactions when it comes to *perceived usefulness*, *clarity of purpose*, *perceived respectfulness* and *perceived emotional safety*. While more positive reactions increase the chance of higher motivation to transfer (Gegenfurtner et al., 2009) this increased motivation can lead to accelerated transfer of training (Burke & Hutchins, 2007), which is a key goal of MD programs.

However, the emerging factor of *becoming aware of assumptions* does not seem as clear-cut when it comes to its relationship with transfer of training. Do test takers who become aware of underlying assumptions see personality tests as instruments of “pigeonholing” (Lundgren, 2012) that may lead to distancing themselves more from the MD program and hence reduce their chances of learning transfer? Or does the critical reflection on the assumptions behind those personality tests lead to an enhanced understanding of the individual in its environment, which would make this MD tool a “learning instrument” and hence increase their learning transfer (Massenberg, Schulte, & Kauffeld, 2017)?

In order to answer this question, it helps to look more closely at concepts of critical reflection (Mezirow, 1998) and the Reflective Judgment Model (P. M. King & Kitchener, 1994), both of which talk about reflection of and on assumptions and link them to learning. According to Mezirow (1998), critical reflection of assumptions is a central concept to understand how adults learn and think for themselves rather than them acting on the concepts, values and feelings of others. The act of reflecting and “turning back” on experience can mean different things to different people; it can come in the form of awareness of assumptions (as shown in the findings), taking something into consideration, “imagining alternatives” (Mezirow, 1998, p. 185) or making an assessment. Critical reflection, on the other hand, involves the questioning of assumptions (eg. “Behavior can be predicted through personality tests”) and presuppositions (eg. “Personality does not change”), which can lead to change in one’s “established frame of reference” (Mezirow, 1998, p. 186). The Reflective Judgment Model (P. M. King & Kitchener, 1994), describes changes in how a learner constructs meaning from assumptions and how these affect the development of critical and reflective thinking. Here, individuals make judgments about controversial issues that are stimulated in the person-environment interaction (Kitchener, King, & DeLuca, 2006).

Theoretical implications

The research findings carry a number of theoretical implications. First of all, by considering six factors that were theoretically derived from selection and feedback literatures, there is first evidence to establish a set of factors that apply to the MD context of the present study. Out of these six factors, some did not seem to show empirical evidence in differentiating test takers’ reactions in MD. Second, additions to existing factors were suggested through first evidence that two emerging factors may expand the model of test takers’ reactions in MD. Third, some dynamics and directionality were identified in one of the emerging factors (*becoming aware of assumptions*) by putting it into the context of existing reflection literatures.

By *becoming aware of assumptions*, test takers might question and challenge personality tests and their presuppositions, however this act of challenging and questioning – while exhibiting negative post-workshop reactions – can lead to an enhanced understanding of what drives and predicts individual behavior. This may lead to a situation where the test takers’ reaction is negative but

the transfer of training is positive; a seemingly paradoxical situation that was already confirmed in other empirical studies on training transfer (Alliger et al., 1997). Hence, test takers may use awareness and reflection of assumptions as a way to enhance and expand their understanding of human behavior and the personality test, especially when discussed controversially and reviewed critically. Here, the personality test becomes a learning instrument.

Although no specific hypotheses were formulated regarding test takers' reactions at the outset of this study, the researchers had an expectation that test takers might be at risk of accepting personality test feedback without questioning it. Also, it was assumed that test takers might not have access to research-based pro's and con's of personality tests and would hence depend completely on the information that the MD facilitator provided to them. However, as the findings show, neither of these two assumptions are true, as test takers feel empowered to decide for themselves how much they take the test feedback for granted as part of their learning paths (Poell & Van der Krogt, 2015). Test takers tend to have their own career development in mind, and when desired they retrieve additional information from external sources, including the Internet (Poell & Van der Krogt, 2014).

Research limitations

Several research limitations to this exploratory study were noted. First of all, despite the small-scale character of this study (N=11), the data collection scope was wide and deep including participatory observations in two different organizations. The study results presented here invite for a follow up research with a larger number of participants, possibly using a quasi-experimental research design.

Secondly, test takers' reactions immediately after the MD workshop were not recorded, despite the acknowledgment in feedback literature that "immediate reactions do matter and they likely influence how the leader responds to the feedback" (Atwater & Brett, 2005, p. 545). In future studies, we suggest that researchers use a survey instrument to collect immediate reactions at the end of the MD workshop and then follow up with interviews three months after. That way, future studies would be able to refer back to those immediate reactions

during the interviews and take them into consideration when linking reactions to motivation to transfer.

Thirdly, a conscious decision was made at the beginning of this research study not to take into consideration actual personality profile data for the qualitative cross-case comparison. Reasons for not collecting actual profile data were the protection of privacy promise that had been given at the beginning of the voluntary sign-up (Anderson, 2017) as well as concerns regarding popular personality tests' reliability and validity, such as the MBTI (Boyle, 1995; Michael, 2003; Pittenger, 2005). However, the researchers later identified feedback literature that shows how individual characteristics of the test taker may influence their responses to and uses of feedback (Atwater et al., 2007). For example, Naquin and Holton (2002) show that personality traits such as extroversion, conscientiousness, neuroticism, agreeableness and openness to experience affect motivation to transfer. Despite aspects of personality traits not being considered in this present study, these personality traits could be a valuable consideration for future research, especially when employing a personality test with high validity and reliability scores, such as the NEO PI-R, a trait-based personality test founded on the Big Five personality theory (Goldberg, 1990; McCrae & Costa, 1999).

Fourthly, the research team did not control for prior experience with personality testing, and the different levels of experience and exposure might have influenced the comparative analysis without making this potential influencing factor explicit.

Suggestions for future research

In order to understand how different factors influence, mediate or moderate test takers' reactions to personality tests, we suggest that researchers conduct a survey study with a large sample in order to test the proportions, directionality and statistical significance of individual, instrument and workshop characteristics as detailed in Figure 5.2 (arrow number one). By conducting this quantitative study, which could also be conducted using a quasi-experimental research design, future studies could confirm this study's exploratory insights, and also distinguish between drivers and their implications for reactions to MD programs where personality tests were used.

Another important area of future research would be a comparative longitudinal study of test takers with positive reactions to personality tests in comparison to those with negative reactions, taking into consideration their immediate reactions after the MD workshop, their personality profile data as well as their actual transfer of training (Figure 5.2, arrow number two). While the first study would investigate further the drivers to test takers' reactions, the second study would confirm the link between reactions and transfer of training in MD workshop settings. Especially with a view on the *becoming aware of assumptions* finding discussed above, the second study would allude to the question how positive and negative reactions to MD workshops in general and personality testing in specific stimulate motivation to transfer and learning, and what effect this motivation has on training transfer and learning outcomes.

A third idea for further research would be to see how our findings could transfer to other practices of MD. For example, some research has been done on the use of 360 degree feedback for development (Atwater & Brett, 2005; Lawrence, 2015; Morgan, Cannan, & Cullinane, 2005), and it would be of interest to find out whether factors such as perceived emotional safety and becoming aware of assumptions also play a role in reactions to these other tools that are often used in MD.

Practical implications

HRD professionals select, administer and discuss the tools that they use during their workshops (Lundgren, Kroon, et al., 2017). When these practitioners choose a personality test for MD, they need to be aware of the factors that influence test takers' reactions in this context. This study shows that workshop content and interaction, such as *clarity of purpose* or *perceived respectfulness*, affect how participants think about the workshop, positive or negative, which could alter their motivation to transfer these insights back into the workplace. HRD professionals need to be aware of their responsibility that they carry in terms of their professional conduct during program facilitation when encouraging a person's learning in "safe spaces" for difficult dialogues (Gayle et al., 2013; Howie & Bagnall, 2015). Practitioners need to consider the question whether their objective is to stimulate positive reactions and acceptance of personality tests, or whether the creation of safe spaces that allow a critical review of assumptions with possible negative reactions as a workshop outcome would be more desirable in the context of MD.

Chapter 6

Discussion and Conclusion



This research illuminates different views of and voices on personality testing when applied in developmental contexts. By adopting a constructivist lens in order to study stakeholders, products, and dynamics this research aims *to explore the practice and perception of personality testing among HRD stakeholders in order to further our understanding of effective and ethical practice in this field.*

The research was conducted in Western Europe, with a focus on personality testing in the UK, Germany and the Netherlands. A number of qualitative methods were employed, combining multiple-case study, qualitative interviews and participatory observations. The data was analyzed using open coding, inductive analysis, qualitative comparative analysis and constant comparison as part of grounded theory. While data was collected between 2012 and 2016, access negotiations were facilitated through the author's prior experiences with personality testing in HRD.

Main findings

At the beginning of this research, a number of research questions and intended contributions were set out. These were structured along the four chapters that form the core of this dissertation.

Chapter 2. How can the personality testing industry be described with its tools, stakeholders and dynamics?

In the opening empirical study, the intended contribution was to highlight personality testing industry dynamics in order to clarify the role of professional organizations, and to make suggestions regarding what it could be. This multiple-case study situated in Western Europe found that industry tensions between psychological associations and non-psychologist practitioners remain unresolved. As a consequence, HRD practitioners find themselves operating in a vacuum, with no professional body to turn to for everyday personality testing questions, or to discuss the ethics of applying these tests in developmental contexts. Psychological associations struggle to find their place between enforcing test standards with their members and/or allowing a wider audience of non-psychologists to administer personality tests in workplace settings. Test publishers have realized this gap and offer free webinars and coaching sessions

to those HRD professionals who are accredited with their tool. This has led to even further fragmentation of the industry, which lacks overall standardization and control over its test use.

HRD professionals who find themselves in this field of unresolved industry tensions where it is not clear what constitutes best practice, have started “re-labeling” their testing practice. Rather than introducing the personality test as “a test”, they talk about tools for personal stocktaking and development. Re-labeling is further employed as a way to demystify or downplay the test while aligning it more with the principles of HRD and adult education; here, formative aspects of “what can be” are prioritized over assessments that confirm “what is”. Considering that each test involves some significant cost components made up of test license and accreditation fees, it is hard to comprehend why this money would be expended for a tool that is then re-labeled as something else in order to make it work for a developmental context. The question comes to mind what alternative methods, approaches or tools could be employed by HRD professionals to fulfill a purpose similar to what personality tests stand for: an exploration of commonalities and differences for individuals or teams, without having to rely on a psychometric test.

In this study, I conclude that HRD professionals who use personality tests need an increased awareness of the dynamics and challenges in the field in order to drive more employee-oriented outcomes. I further recommend that HRD associations look critically at this trend where commercial publishers give guidance to practitioners. HRD associations should also reconsider what their own role can be in the personality testing ecosystem, given their purpose and access to practitioners.

Chapter 3. How has existing literature conceptualized and operationalized critical reflection?

The intended contribution to the conceptual part of this dissertation research was to develop a methodology based on empirical critical reflection research in order to identify practical ways of studying reflection in the workplace. This literature review concludes that there is little agreement in adult education on how to operationalize reflection, because existing studies differ significantly

in terms of their aims, research designs, data collection methods and settings in which they are conducted. For example, one study aimed to develop a new coding scheme (Kember et al., 1999), while another study evaluated an existing scheme when applying it to a different setting (Wallman et al., 2008). Especially in medical education studies, the emphasis was on evaluating students' critical reflection skills (Liimatainen, Poskiparta, Karhila, & Sjogren, 2001), while the aspect of evaluation was less important in studies within professional development (Peltier et al., 2005). In conclusion of this conceptual paper, I therefore suggest possible improvements when operationalizing reflection in empirical HRD studies, namely to integrate different critical reflection traditions, to use multiple data collection pathways, to opt for thematic embedding, and to attend to feelings (Lundgren & Poell, 2016).

These four improvements derived from literature were used as guiding principles when conducting the remaining two empirical studies on HRD professionals' strategies and test takers' reactions, and when analyzing the collected data. Building on Mezirow's (1978, 1981, 1990, 1991, 1998) critical reflection theory and hence approaching the research from a constructivist perspective worked for some parts of the consecutive empirical studies, but also provided its challenges. For example, HRD professionals were asked to reflect on their practice in an interview, after the fact and separate from their natural work setting, e.g., a classroom workshop, distant from their personality tests used, and away from internal and external stakeholders, e.g., the test takers, publishers or colleagues. I was not able to observe the professionals' actual interaction with the tool itself within the ecosystem of products, stakeholders and industry dynamics that all contribute to their reflection process and outcomes. In that sense, the suggested critical reflection methodology was only partially applied in the empirical studies of this dissertation.

Going forward, the operationalized reflection framework could be enhanced by combining different perspectives on reflection and learning (Lundgren, Bang, et al., 2017) to complement the dominant constructivist approach used in this research. For example, by adding a situative perspective (Justice, Lundgren, Marsick, Poell, & Yorks, 2018), a richer picture of HRD practitioners and their interactions with tests as tools of professional practice could be achieved. By implementing these improvements, the projection is enhanced critical reflection

research in order to understand more fully practitioners' choices and strategies of tool use.

Chapter 4. How do HRD professionals make sense of their own personality testing practice?

For the qualitative empirical study with HRD professionals, the intended contribution was to explore how experienced coaches, business trainers and HRD managers approach personality testing in their own practice. The intent was to develop guidelines for newer practitioners who are entering the personality testing field. The study summarizes reflections of 18 professionals who were interviewed as part of this inductive analysis. The findings show that HRD professionals apply six strategies when using personality tests in their practice. Here, the spectrum ranges from more protective and selective strategies to approaches that can be described as pragmatic and accommodating, while some practitioners decide to avoid personality tests because of ethical concerns.

The theory of cognitive dissonance helps in part to explain how practitioners move from one approach to another depending on their cognition and critical awareness of tests as tools of their professional practice in the workplace. This critical awareness is modulated by organizational sensemaking in the context of the entities that these professional interacted with. The study further shows that professionals usually do not start from a "clean slate" but are confronted with the preferred tests already acquired by the organization (relevant for internal HRD professionals) or with desired tests that are requested by the client organization (relevant for external HRD coaches, trainers or psychologists). Here again, the environment and ecosystem of products, stakeholders and industry dynamics that an HRD professional deals with seem to influence and possibly determine the professionals' strategy.

The intended contribution was to explore the HRD professionals' approaches; in that sense, the first part of this intent has been achieved through this study. However, a couple of questions remain unanswered in this context. For example, whether personality test accreditation is seen by employers and clients as qualification of these HRD professionals and hence increases their market value is still unclear. Furthermore, more inquiry is needed to establish whether consul-

tants opt for specific personality tests as a unique selling point. For example, one could hypothesize that those consultancies only using trait-tests could be perceived as more mature, scientific and upscale in comparison to type-test using consultancies. Lastly, it remains to be seen whether tests that pay attention to individuals are more appreciated in individualistic cultures (e.g., Germany, UK, USA) compared to collectivist cultures (e.g., China, Korea, Vietnam) where the application of such a tool might be less desirable. A number of these remaining questions would need to be explored further before more universal guidance to newer practitioners can be given.

Chapter 5. How do test takers react to personality testing in developmental contexts?

For the last empirical study the intended contribution was to establish a framework of test takers' perceptions of personality testing in HRD in order to advance the literature in developmental contexts. Here, influencing factors that had been derived from management selection studies were used as conditions to compare different test takers' reactions in a longitudinal study with eleven participants of two different management development programs.

This study finds that test takers reflect more critically on the use of personality testing in management development than previously assumed in management selection literatures (Beermann et al., 2013). In this context, the HRD professional when acting as a workshop facilitator, trainer or coach plays a big role in how test takers react to personality testing: by clarifying the purpose, showing its usefulness and treating test takers with respect the chances of positive reactions increase. The results further show that the interaction between test taker and HRD professional influences the test takers' reaction to a larger extent than the tool itself. In addition to *perceived usefulness*, *clarity of purpose*, *perceived respectfulness* – three factors that had already been established in management selection – *becoming aware of assumptions* and *perceived emotional safety* that is experienced during the workshop, both influence test takers' reactions in management development. Test takers who actively engage in questioning their assumptions and who feel emotionally safe in the setting are more likely to be positive about personality test use, regardless of which test was used.

That said, HRD professionals can influence all of the five factors directly through their actions, words and behaviors during the workshop interaction. Therefore, organizations that allow personality testing in developmental contexts need to make sure that facilitators are aware on their impact on test takers' reactions. What is meant as a learning instrument for growth and advancement can easily be perceived as a mechanism of "pigeonholing". As long as professional associations in the area of personnel management and corporate learning, such as the professional body for HR and people development CIPD in the UK, do not strengthen their presence and involvement with HRD professionals, hosting organizations should decide how they want to train up their facilitators to engage with test takers in an ethical way using "positive practice" (Geue, 2018) in the workplace. This finding links with the conclusion of Chapter 2 that requests HRD associations to reconsider their own role in the personality testing ecosystem.

Overarching contribution

The overarching intended contribution of this research was to encourage critical reflection of and enhanced discourse on the ethics of test use for improved professional practice in HRD. This aspect ties in closely with the understanding of professionalism among HRD practitioners. Professionalism can be defined as an altruistic service to the public, practiced with ethical diligence, integrity and guarantee of quality (Sullivan, 2005). Professionals, such as HRD practitioners, are awarded some autonomy in their practice, status of its knowledge, financial reward and self-regulation in exchange for the guarantee of professionalism. This has led to requirements for professional education to inject moral attitudes and behaviors in new professionals that will provide this service. In HRD, professional education has been offered through higher and continuing education (Poell, 2015; Stewart, Mills, & Sambrook, 2015), as well as through accreditation courses presented by test publishers themselves. This poses a conflict of interest, as test publishers have a commercial stake that does not align smoothly with an altruistic service to the public. In addition, and as Fenwick (2016) notes, the requirement for professional education has led to the tendency to define professionalism as a set of ethical behaviors rather than as a more critical orientation on the practice itself. One outcome of Chapter 4 that discusses strategies of HRD professionals was that practitioners are largely divided when it comes to the ethical question of "what is good and what is bad practice"? Asking this

question is a starting point, but at the same time, it does not encourage HRD professionals to assess further the dynamics and struggles in the field of personality testing. As such, critical reflection paired with a discourse among scholars and practitioners could be beneficial to refocus the efforts around shaping ethical practice and a shared code of conduct.

This research has already triggered a number of informal discussions that happened during access negotiations, in interview settings, through workshops, in conference sessions and during research group meetings along the way. However, the findings of this study, especially concerning the strategies used by HRD professionals in their practice, have not yet been formatted to be used for wider discourse on professionalism when it comes to personality test use in HRD. The challenge going forward will be to translate the present research findings into didactic content that can be used in higher and professional education to help newer practitioners to make critically reflective choices when they start their personality testing practice.

Reflections on theories

Completing a research study provides an opportunity to pause and think about one's own scholarly practice in the process of providing reflections on theories used in this research. Three reflections on theory enhancements are offered below.

Mezirow's reflection framework as a conceptual support structure as laid out in Chapter 3 proved useful as a starting point, but was limiting in scope when trying to operationalize it for the two qualitative empirical studies presented in this research (Chapters 4 and 5). Since Mezirow included mostly cognitive components in his conceptual frame where reflection is constructed in the head of the study's participant, insufficient materials that captured the dynamics surrounding the individual were included in this framework. Here, the addition of a situative perspective (cf. Brown et al., 1989; Greeno, 1997; Sfard, 1998) could add value in order to study more closely the interaction with the personality test as tool in itself, as well as the interaction among HRD practitioners and/or between practitioners and other stakeholders in the testing industry. Viewed from a situative perspective, reflection is one of the many learning processes that happen in

the workplace when employees interact with each other and through tools and systems (Lundgren, Bang, et al., 2017). While the situative perspective focuses on participation and collaboration between organizational stakeholders and their environment (Lundin & Nuldén, 2007), tools such as a personality test can trigger reflection and discussion within the workplace. Hence, a hybrid constructivist-situative framework would be potentially feasible (Justice et al., 2018), with the possibility to enhance the reflection framework even further with psycho-analytical, enactivist and critical-cultural perspectives (Lundgren, Bang, et al., 2017). To tie this line of thinking back to personality testing, HRM theories of test use need adapting while tools shift from personnel selection to MD areas of HRD. With this shift in the environment, the question of adequate frameworks and approaches needs to be discussed. Parallels can be drawn with the shift in career counseling that can be described by situative approaches such as the Chaos Theory of Careers (Pryor & Bright, 2014) and the “happenstance” learning theory (Krumboltz, 2009). The focus on the individual in isolation is not sufficient to inform about the dynamics and interactions that contribute to personality testing in HRD. Researchers need to adopt multiple frameworks and develop data collection techniques that operationalize these perspectives.

The six strategies of HRD professionals’ test use presented in Chapter 4 are rather narrowly informed. Consulting the theory of cognitive dissonance (Festinger, 1962) as a proxy to explain the observed cognition-behavior gap among HRD practitioners, and how some have shifted from one strategy in applying personality tests to another, had face value at first. However, reflecting on it more, intra-personal cognitive dissonance is not sufficient to explain all facets of why an HRD professional is motivated to employ one strategy and not another, or what motivated an HRD professional to shift their strategy as they matured in their practice. While cognitive dissonance theory is often applied in management studies (Hinojosa, Gardner, Walker, Cogliser, & Gullifor, 2017), it leaves out organizational factors, such as the impact of line management directions, organizational history or industry-specific dynamics that will have influenced the HRD professionals’ motivation and decisions. Similar to the comment made above on Mezirow’s framework, cognitive dissonance has its focus more on the individual’s meaning making, and less on organizational sensemaking. Realizing this limitation after the data analysis of Chapter 4 was done, I used the “paradox in organizations” (Lewis & Smith, 2014) as an alternative lens, without exploiting

that perspective fully in this dissertation. Also, the study of “institutional logics” (Besharov & Smith, 2014) that describe socially-constructed sets of “material practices, assumptions, values, and beliefs” (p. 364) could be complimentary to explain an organization’s impact on an individual’s cognition and behavior.

Lastly, one way to generalize these findings and to turn from personality tests in specific to learning tools in general, would be to link this study to the ideas of sociomaterialism (Fenwick, 2014, 2016; Fenwick & Edwards, 2017). A sociomaterial perspective depicts materials as “dynamic and enmeshed with human activity in everyday practices” (Fenwick, 2014, p. 47). These materials can be objects, technologies or other artifacts of workplace life, which would include the personality test as a material object that HRD professionals and test users are dynamically entangled with. As much as there is a materials component in terms of selecting the test, distributing the link, administrating test taker submissions and printing the feedback pages, there is also a social dynamic in the classroom when HRD professionals “negotiate” the purpose of the learning activity and re-label the test as a tool for stocktaking and self-discovery. In a similar way, other tools frequently used in HRD, such as 360-degree feedback tools, could be analyzed from a sociomaterial perspective.

Reflections on methods

In the following section, three reflections on the use of methods in this research will be discussed.

Firstly, studying a field through a multiple-case study means employing a wide and diffused approach. As Chapter 3 draws on many different ethnographic fieldwork-type sources including interviews, test reports, product flyers, email correspondence and others, this approach facilitated depicting a broad industry overview including its internal and external stakeholders, tools, and major themes of tension. However, the actual interaction between publishers and associations or psychologists and non-psychologists was not observed, rather it was constructed through triangulation from the data collected. To complete the picture, a series of focus group discussions with members of opposing stakeholders could add depth to understanding the nature of the conflict, and these focus group discussions could provide grounds for a resolution or egress from

this tension. In addition, the choice for Chapter 3 to study stakeholders in three different geographies (UK, Germany, The Netherlands) worked well for a comparative analysis in the form of the present multiple-case study analysis, but no direct interactions between these different geographic stakeholders were recorded. Here, an exchange of ideas could provide useful in terms of learning from each other's best practice to navigate this industry. Further questions for related explorations include: How do HRD practitioners make sense of the differences that exist across Western Europe? Do the findings of this multiple-case study apply to other parts of Europe, e.g., Southern Europe and/or other regions of the world, e.g., Asia? This research was based in Western Europe and conducted by scholars who have lived and worked mostly in Western Europe. Therefore, broadening the geographic scope could provoke new discussions and untapped perspectives.

In retrospect, collecting data for the HRD professionals study (Chapter 4) in the form of interviews seems one-dimensional. The point of view adopted here can be described as an "after-the-fact" perspective as HRD professionals were asked to reflect on their experience, values and attitudes in relation to personality testing. The reflection was *on action* rather than *in action* (Schön, 1983), and it can be assumed that some practitioners formulated their thoughts and ideas during the question and answer pattern of the research interview rather than *a priori*. Hence, the way the questions were asked and the way that the interview between HRD professionals and researcher unfolded will have influenced the how practitioners positioned themselves and their test use strategies as a product of our conversations. Therefore, a longitudinal study design of participatory observations in addition to interviews with HRD professionals could have been useful to collect *in action* and *on action* data points, similar to the research design of the test taker study (Chapter 5). Noticing and recording those "micro-practices" (Rouleau, 2005) of HRD professionals as they facilitate workshops or team interventions and inquiring afterwards about those observed practices, could have enriched the empirical study described in Chapter 4. This would have also helped to more clearly differentiate between organizational *sensemaking* (Rouleau, 2005, p. 1415), that is, the way HRD professionals understand, interpret, and create sense for themselves based on the information they retrieve from their work environment and *sensegiving* (Rouleau, 2005, p. 1415), that is, how HRD professionals attempt to influence test takers' reactions, communicate

their thoughts about the personality test at hand, and gain stakeholder support for their approach. While a longitudinal study with participatory observations poses challenges in terms of accessing sites for research and getting permission to “observe” sensitive things like personality test feedback sessions based on instrumentation, these challenges could be less pronounced among social media-savvy employees who might be less sensitive of having their “souls bared” in front of strangers.

Thirdly, no individual personality profile data was collected. In none of the three empirical studies (Chapters 3, 4 and 5) did the research design incorporate any individual personality profile data to feed into the formulation of HRD professionals’ strategies or the test takers’ reactions. Now, at the conclusion of this research, the question arises whether this was a missed strategy or a strategic omission aligned with the overall learning instrument positioning of the three studies. Although feedback literatures (cf. Atwater et al., 2007) discuss how test takers’ reactions are influenced by their individual characteristics, the present studies did not operationalize this insight into the research design in Chapter 5. One reason for neglecting personality profile data as input to this study was that many of the test takers were confronted with type-based personality tests including validity and reliability concerns. These type-based tools would have been inappropriate to use as data points for the research. However, the researcher could have asked test takers to also complete a short version of the NEO PI-R questionnaire (based on the Big Five/Five Factor Model) that is commonly used in research studies, and then triangulate Chapter 5’s findings with the personality profile results of each test taker. Reflecting on this question now, after its strategic omission, the concern persists that using a personality test to study the perception of personality testing would have pushed the whole research project into a less desirable direction. On the one hand, it might have made access negotiations more difficult because of different approval levels that are required to use a psychometric instrument as part of the research design. On the other hand, the willingness to participate in this study, especially for those HRD professionals and test takers with concerns or negative perceptions, might have been lowered had participants been asked to complete an additional personality test before the interviews. After all, the empirical studies presented in Chapters 4 and 5 were “not a test” but an exploration of individual reflections and meaning-making structures.

Implications for future research

A number of future research streams could result from this study. Combing the most pressing questions into a distinct number of research projects could result into the following four:

Future research theme 1: Learning tools and paths

The outcomes of this dissertation suggest the adoption of a constructivist-situative perspective, in which reflection and learning emerges in the interaction between individuals within their (professional) environment, including the use of tools. Work-based tools and systems can trigger collaboration between learners that requires minimum organizing or formal intervention. What is the role of personality tests and other tools in individual learning paths? Going forward, I plan to continue research on learning tools in the workplace, while connecting to the larger research theme of learning paths (Poell et al., 2018; Poell & Van der Krogt, 2010) within Learning-Network Theory (LNT, cf. Poell & Van Der Krogt, 2017a, 2017b; Poell & Van der Krogt, 2015).

Future research theme 2: Social media enabled personality profiling

Self-assessments in personality testing are likely to soon become a tool of the past. Studies show how digital records of Facebook online behavior (Kosinski, Stillwell, & Graepel, 2013; Park, 2010; Schwartz et al., 2013), Twitter messages (Carducci, Rizzo, Monti, Palumbo, & Morisio, 2018) and LinkedIn profile data (van de Ven, Bogaert, Serlie, Brandt, & Denissen, 2017) provide convenient and reliable ways to measure psychological traits. Even computer predictions based on a generic digital footprint, such as Facebook likes, claim to be more accurate than predictions made by the study participants' friends (Youyou, Kosinski, & Stillwell, 2015). These developments that shift the industry away from self-report assessments and toward predictions based on digital records of behavior raise a number of questions about the practice and ethics of social media enabled personality profiling. What is the future of self-assessment personality tests in light of social media enabled personality profiling? Will one replace the other, or will both forms co-exist and complement one another? What role should HR play in enabling or discouraging decision-making based on digital records-based profiling in times when organizations are striving to reduce diversity resistance (Wiggins-Romesburg & Githens, 2018) in order to create more diverse

and inclusive work environments (Bourke, Garr, van Berkel, & Wong, 2017)? This future research theme could start with an ethnographic field study that describes current practice of social media enabled personality profiling in organizations. Potentially, the study's findings could be used to triangulate the conclusions of Chapter 2.

Future research theme 3: Reflection at work

How to operationalize reflection when adopting multiple perspectives? What are the boundaries of reflection? This research theme is a continuation of the many individual meaning-making and organizational sensemaking questions asked about reflection as an action that affects practice. By moving reflection research from "inside the head" (constructivism) to other locations and learning ecologies, new connections and interactions with psychoanalytic (Enckell, 2010), critical-cultural (Fenwick, 2000b; Giroux, 2005) and enactivist perspectives (Justice & Yorks, 2018) become possible. In this context, the ethnographic research around sociomaterialism (Fenwick, 2016; Fenwick & Edwards, 2017) should be considered as an extension where the entanglement of material objects, their users and communication between the two could be studied.

Future research theme 4: Transformation of HRD

As a scholar-practitioner, I notice that HRD is no longer expected to be the primary agency for promoting learning and development along tools such as personality tests among employees. Today, HRD is diffused and integrated into a broad range of leadership and supervisory roles (Torraco & Lundgren, 2018), where multi-functional individuals at different levels of the organization introduce individual and team development tools without necessarily consulting with the HR department. Previous "state-of-the-art" studies have looked at generational differences in the workplace (Deal, Altman, & Rogelberg, 2010; Lyons & Kuron, 2014), the role of technology in HRD (Church, Gilbert, Oliver, Paquet, & Surface, 2002; Thomas, 2014), and the impact that informal and self-directed learning have on the profession (Ellinger, 2004; Nesbit, 2012). Hence, as more responsibility for learning and development is assumed by others and automated by technology, what role should HRD professionals play as gatekeepers of tool use in organizations? And who are the stakeholders that define the boundaries and best practices in T&D tool use going forward? These questions on the role of HRD in this context could be addressed in a series of empirical studies that

review how HRD practitioners experience the adaptations and advances in tool use within different organizational settings, with the potential of rolling out a transnational research study.

Implications for practice

From the discussion and reflections in this closing chapter, I derive a number of implications for practice.

Ethical concerns are prevalent among critically reflective test takers as shown in Chapter 5. However, these concerns were not at the forefront of the majority of HRD practitioners who addressed *perceived respectfulness* and *perceived emotional safety* more as an afterthought than as a priority concern. Given this disconnect in perspectives and priorities, test takers more easily feel “pigeonholed” while HRD professionals strive to re-label the tool as a “learning instrument”. One implication for practice is to discuss this communicative disconnect more openly. For example, HRD professionals need to address issues of pigeonholing risks associated with personality testing as part of their workshop delivery, rather than leaving this decision to the test takers.

Secondly, the HRD professionals study presented in Chapter 4 hinted towards various demographic factors that influence test practice including academic degree and years of professional experience. The higher the academic degree and the longer someone has worked with personality tests, the more likely this HRD professional will use an *ethical-protective* or a *scientific-selective* strategy. But what about those ‘novice’ practitioners who are just starting out? Who will give them guidance, other than commercially oriented test publishers? In this context, the role of professional learning and development associations needs to be clarified and possibly strengthened as this relates to professionalism and professionalization in HRD and the question of what makes a good professional in this field.

Finally, major concerns persist about industry regulation as well as the upholding of good test practice standards, as guidance to novice practitioners is mainly given by commercially oriented test publishers. In this context an implication for practice is to inquire more deeply about the role that HRD plays as a gatekeeping

function. Also, there is renewed interest in personality profile data since the Cambridge Analytica case (Chen & Potenza, 2018): Who will regulate the use of profile data on the transnational level? What portion of this will the HRD function assume responsibility for when it comes to the collection, aggregation and dissemination of employee data for profiling?

Conclusion: Pigeonholing and learning instrument

Personality tests can be used as tools to stereotype, pigeonhole, and departmentalize. When employed as such a tool, the test feedback emphasizes difference and disparity that can be exploited as justification for ill-suited behavior in the workplace. This is especially true when personality tests are introduced as tools of absolute truth – a gospel that leaves little room for questioning assumptions. However, concurrently, personality tests can also be enacted as tools to learn, question and grow. When introduced and handled that way, the test feedback emphasizes commonalities and relationships that can enable more inclusive work environments. This is especially true when the distinction is made between personality and behavior, decoupling the two and seeking to identify where opportunities for improved workplace behaviors lie. Used as a learning tool, growth and development of individual learners can be supported through personality tests.

HRD professionals as tool facilitators play a major role in the pigeonholing versus learning instrument debate. Here, it comes down to the approach and strategy employed to introduce, administer and implement the test and its feedback. A practitioner pointed it out early in the research when he referred to a “going to the dentist” analogy (Doc020, lines 192-93). This HRD professional described how, as a patient, he is interested in the quality of the dentist’s practice rather than the quality of the instruments this dentist might be using. Following this analogy, the dichotomy of “pigeonholing or learning instrument” is problematic as it might not be the quality of the test but the quality of the tester that we should pay attention to. The focus has shifted from tool to tool handler.

The dichotomy is also problematic as the condition of work has shifted: from a more stable employment where organizations hired for life to a work environment where contingent and freelance work have become the norm. Reactions

to personality tests are situated in its work environment, its power structures and organizational climate. With this context in mind, it is understandable that some test takers are more critical towards personality tests and possibly more suspicious to what extent these tools might be used to their disadvantage. From this point of view, workplace cultures have become less trusting and less “safe” places. At the same time, younger employees have become accustomed to standardized testing as part of their education. These younger test takers might also find it less problematic to submit personal information in these personality questionnaires as they are used to doing so in online marketplaces as well as through social media.

This dissertation began with two quotes. Sophie, a young education specialist who felt pigeonholed by her manager and decided not to apply personality testing in her own professional practice. Jackie, on the other hand, a more seasoned learning and development manager who recognized the risk of pigeonholing and hence decided to introduce the test using softer terms and a more developmental approach. How could Sophie’s experience have been different had she been in a different work context with exposure to a practitioner like Jackie? Sophie might have felt more at ease and more trusting with Jackie’s approach. Sophie might also have been encouraged to reflect on the pigeonholing challenges that come with such a tool. Finally, Sophie might have seen how personality testing can be pigeonholing and learning instrument at the same time.

Summary

Summary

Personality tests like the Myers–Briggs type indicator (MBTI) or Big Five assessments enjoy great popularity in business, human resource development (HRD), and adult education settings. Psychometrics and personality testing are fascinating fields of study, with numerous writings on its history, its development, and its role in the workplace. Originally, personality tests were designed for personnel selection. Increasingly, they are also used in developmental areas of the human resources spectrum, such as coaching, educational leadership, organizational change, team building, and management development.

It is estimated that four out of ten tests are used in team building and management development, often referred to as tools for self-exploration and self-reflection, and with the aim to improve team performance. The market size is considerable: about 2,500 different tests are administered a few million times every year, producing an estimated turnover of US\$500m to US\$2bn per annum. As new technologies support the quick creation and dissemination of tests, it can be assumed that the personality test market is growing even further. This means that employees encounter with great certainty a moment in their career when their personality is assessed in organizational life.

However, while personality tests are often promoted as enablers to enhance personal development and organizational functioning, they can also yield the opposite effect of departmentalization and stereotyping. Criticized in many ways, a number of frequently used personality tests have poor validity and show methodological issues around forced-choice answer formats. Another contradiction evolves around the static nature of personality itself, which leaves scanty room for change and development. HRD at its core is about seeking opportunities for learning, growth, and development. The question thus arises how a test that determines “what is” can be effectively used in an adult education setting that strives to find out “what can be”.

While the use of personality tests remains a controversial topic, very little literature exists that explains their use in developmental contexts, such as management development. Without a clear understanding of personality testing in settings other than selection, the current practice raises ethical concerns about

the rights and responsibilities of test takers. For example, test takers might be given the impression that they are being assessed, even if that is not the aim in a workplace training setting. Also, test takers might feel uncomfortable sharing test feedback at work with their line managers and colleagues.

This dissertation aims to explore the practice and perception of personality testing in HRD in order to build on ethical and effective practice in the field.

The research was conducted in Western Europe, with a focus on personality testing in the UK, Germany, and the Netherlands. With the exception of a few international HRD practitioners from Denmark and the United States, all study participants were based in one of these Western European countries. A set of qualitative methods was employed, combining multiple-case study, qualitative interviews, and participatory observations. Resulting data was analyzed using open coding, inductive analysis, qualitative comparative analysis, and constant comparison. During the analysis, special attention was given to the role of critical reflection as an indicator for deliberate professional practice with regard to the purpose of HRD practitioners and test takers.

This dissertation is structured around four studies that each have distinct research objectives.

- *Study 1* explores the personality testing industry. This study inquires about the role of internal and external stakeholders, the value of psychometric and practical considerations in test selection, and the purpose of personality test use in workplace training.
- The focus of *Study 2* – in the form of a literature review – is on critical reflection. This study reviews existing reflection research and how its operationalization informs the current and future research.
- *Study 3* focuses on HRD professionals and their personality testing practice. This study illustrates individual meaning-making and organizational sense-making structures of practitioners.

-
- *Study 4* concerns test takers who participate in management development workshops where a personality test is used. This study examines differences in test takers' reactions and factors that influence these.

Study 1: How can the personality testing industry be described with its instruments, stakeholders and dynamics?

The contribution of the first study is to highlight personality testing industry dynamics in order to clarify the role of professional organizations, and to make suggestions regarding what it could be. This multiple-case study uncovers that industry tensions between psychological associations and non-psychologist practitioners remain unresolved. As a consequence, HRD practitioners find themselves operating in a vacuum, with no professional body to turn to for everyday personality testing questions, or to discuss the ethics of applying these tests in developmental contexts. Psychological associations struggle to find their place between enforcing test standards with their members and/or allowing a wider audience of non-psychologists to administer personality tests in workplace settings. Test publishers have realized this gap and offer free webinars and coaching sessions to those HRD professionals who are accredited with their tool. This has led to even further fragmentation of the industry, which lacks overall standardization and control over its test use.

HRD professionals who find themselves in this field of unresolved industry tensions, where it is not clear what constitutes best practice, have started "re-labeling" their testing practice. Rather than introducing the personality test as "a test", they talk about tools for personal stocktaking and development. Re-labeling is further employed as a way to demystify or downplay the test while aligning it more with the principles of HRD and adult education; here, formative aspects of "what can be" are prioritized over assessments that confirm "what is". Considering that each test involves some significant cost components made up of test license and accreditation fees, it is hard to comprehend why this money would be expended for a tool that is then re-labeled as something else in order to make it work for a developmental context.

Study 1 concludes that HRD professionals who use personality tests need to sharpen their awareness of the dynamics and challenges in the field in order to

drive more employee-oriented outcomes, and it recommends that HRD associations look critically at this trend where commercial publishers give guidance to practitioners. HRD associations should also reconsider what their own role can be in the personality testing ecosystem, given their purpose and access to practitioners.

Study 2: How has existing literature conceptualized and operationalized critical reflection?

The contribution of this literature review is to develop a methodology based on empirical critical reflection research in order to identify practical ways of studying reflection in the workplace. The concepts of content, process, and premise reflection have often been cited, and operationalizing Mezirow's (1978, 1981, 1990, 1991, 1998) high-level transformative learning theory and its components has been the endeavor of adult education and HRD researchers¹. By conducting a literature review of 12 empirical studies on critical reflection in contexts of adult education, *Study 2* finds that there is little agreement on how to operationalize reflection. Existing studies differ significantly in terms of their aims, research designs, data collection methods and settings in which they are conducted. For example, one study aimed to develop a new coding scheme while another study evaluated an existing scheme when applying it to a different setting. Especially in medical education studies, the emphasis was on evaluating students' critical reflection skills, while the aspect of evaluation was less important in studies within professional and management development.

Study 2 suggests four possible improvements when operationalizing reflection in empirical HRD studies, namely to integrate different critical reflection traditions, to use multiple data collection pathways, to opt for thematic embedding, and to attend to feelings. These four improvements were used as guiding principles when conducting the remaining two empirical studies on HRD professionals' strategies and test takers' reactions, and when analyzing the collected data.

¹ Jack Mezirow (1923-2014), known for his Transformative Learning Theory, was an American sociologist and emeritus professor of adult and continuing education at Teachers College, Columbia University.

Study 3: How do HRD professionals make sense of their own personality testing practice?

This qualitative empirical study explores how experienced coaches, business trainers and HRD managers approach personality testing in their own practice, with the intent to develop guidelines for newer practitioners who are entering the personality testing field. The study summarizes reflections of 18 professionals who were interviewed as part of this analysis. The findings show that HRD professionals apply six strategies when using personality tests in their practice:

1. *Ethical–protective*
2. *Scientific–selective*
3. *Cautious–avoiding*
4. *Cautious–embracing*
5. *User friendly–pragmatic*
6. *Knowledgeable–accommodating*

Here, the spectrum ranges from more protective and selective strategies to approaches that can be described as pragmatic and accommodating, while some practitioners decide to avoid personality tests altogether because of ethical concerns.

The theory of cognitive dissonance² helps in part to explain how practitioners move from one strategy to another, depending on their perception and critical awareness of personality tests and their application in developmental contexts. This critical awareness is modulated by organizational sensemaking in the context of the entities that these professional interacted with. The study further shows that professionals usually do not start from a clean slate but are confronted with the preferred tests already acquired by the organization or with desired tests that are requested by the client organization. Here again, the environment and ecosystem of products, stakeholders, and industry dynamics that an HRD professional deals with seems to influence and possibly determine the professionals' strategy.

² For an elaboration on this theory, see Festinger, L. (1962). *A theory of cognitive dissonance* (Vol. 2). Palo Alto, CA: Stanford University Press.

A couple of questions remain unanswered in *Study 3*. For example, whether personality test accreditation is seen by employers and clients as qualification of these HRD professionals and hence increases their market value is still unclear. Furthermore, more inquiry is needed to establish whether consultants opt for specific personality tests as a unique selling point. One could hypothesize that those consultancies only using trait-based tests could be perceived as more mature, scientific, and upscale in comparison to type-based test using consultancies. Lastly, it remains to be seen whether tests are perceived differently across national cultures.

Study 4: How do test takers react to personality testing in developmental contexts?

The final empirical study aims to establish a framework of test takers' perceptions of personality testing in HRD in order to advance the literature in developmental contexts. Here, influencing factors that had been derived from management selection studies were used as conditions to compare different test takers' reactions in a longitudinal approach with eleven participants of two different management development programs.

The study finds that test takers reflect more critically on the use of personality testing in management development than previously assumed in management selection literatures. The results further show that the interaction between test taker and HRD professional influences the test takers' reaction to a larger extent than the tool itself: by addressing the purpose of the test use (*clarity of purpose*), discussing its utility (*perceived usefulness*), and treating test takers in a civil and polite way (*perceived respectfulness*), the chance of positive test taker reactions increases. In addition, *becoming aware of assumptions* and the *perceived emotional safety* that is experienced during the workshop, both influence test takers' reactions in management development. *Study 4* finds that test takers who actively engage in questioning their assumptions and who feel emotionally safe in the setting are more likely to be positive about personality test use, regardless of which personality test was used.

That said, HRD professionals can influence all of the above mentioned five factors directly through their actions and behaviors during the workshop in-

teraction. Therefore, organizations that allow personality testing in developmental contexts need to make sure that facilitators are aware on their impact on test takers' reactions. What is meant as a learning instrument for growth and advancement can easily be perceived as a mechanism of "pigeonholing". As long as professional associations in the area of personnel management and corporate learning do not strengthen their presence and involvement with HRD professionals, hosting organizations should decide how they want to train up their facilitators to engage with test takers in an ethical way using positive practice in the workplace. This finding links with the conclusion of *Study 1* that challenges HRD associations to reconsider their own role in the personality-testing ecosystem.

Overall contributions

This dissertation illuminates different perceptions of personality testing when applied in developmental contexts. By studying stakeholders, products, and dynamics this research explored the practice and perception of personality testing in HRD with the aim to further our understanding of strategies and approaches in this field.

One overall contribution of this research is to encourage critical reflection of and enhanced discourse on the ethics of test use for improved professional practice in HRD. This aspect ties in closely with the understanding of professionalism among HRD practitioners who are awarded a fair amount of autonomy in their practice. Test publishers play a major role in the personality testing industry, whether it is through their accreditation programs or through their consultative selling practices. This poses a conflict of interest, as test publishers have a commercial interest that does not align smoothly with an altruistic service to the public. As described in *Study 3*, practitioners are largely divided when it comes to the ethical question of what is good/bad practice. Asking this question is a starting point, but at the same time it does not encourage HRD professionals to assess further the dynamics and struggles in the field of personality testing. As such, critical reflection paired with a discourse among scholars and practitioners could be beneficial to refocus the efforts around shaping ethical practice and a shared code of conduct. This research has already triggered a number of informal discussions that happened during access negotiations, in interview

settings, through workshops, in conference sessions and during research group meetings along the way. However, the findings of this study, especially concerning the strategies used by HRD professionals in their practice, have not yet been presented in a way that it can be used for wider discourse on professionalism when it comes to personality test use in HRD. The challenge going forward will be to translate the dissertation findings into didactic content that can be used in higher and professional education to help HRD professionals to make critically reflective choices in their personality testing practice.

Pigeonholing or learning instrument?

Personality tests can be used as tools to stereotype, pigeonhole, and departmentalize. When employed as such a tool, the test feedback emphasizes difference and disparity that can be exploited as justification for ill-suited behavior in the workplace. This is especially true when personality tests are introduced as tools of absolute truth – a gospel that leaves little room for questioning assumptions. However, concurrently, personality tests can also be enacted as tools to learn, question and grow. When introduced and handled that way, the test feedback emphasizes commonalities and relationships that can enable more inclusive work environments. This is especially true when the distinction is made between personality and behavior, decoupling the two and seeking to identify where opportunities for improved workplace behaviors lie. Used as a learning tool, growth, and development of individual learners can be supported through personality tests.

HRD professionals as facilitators of personality tests play a major role in the pigeonholing-versus-learning instrument debate. Here, it comes down to the approach and strategy employed to introduce, administer, and implement the test and its feedback. A practitioner in *Study 1* described this insight early in the research when he compared the experience of taking tests with going to the dentist. This HRD professional explained how, as a patient, he was interested in the quality of the dentist's practice rather than the quality of the instruments the dentist might be using. Following this analogy, the dichotomy of "pigeonholing versus learning instrument" is problematic as it might not be the quality of the test but the quality of the facilitator that we should pay attention to.

The dichotomy is also problematic, as the condition of work has shifted from more stable employment where organizations hired for life to a work environment where contingent and freelance work have become the norm. Reactions to personality tests are situated in their work environment, their power structures and organizational climate. With this context in mind, it is understandable that some test takers might be more critical towards personality tests and how they are employed in developmental contexts.

Zusammenfassung

Zusammenfassung

Persönlichkeitstests wie der Myers-Briggs-Typenindikator (MBTI) oder das Big Five Persönlichkeits-Bewertungssystem sind Instrumente, die sich großer Beliebtheit in der Wirtschaft, der Personalentwicklung (PE) und der Erwachsenenbildung erfreuen. Mit zahlreichen Veröffentlichungen über ihre Geschichte, ihre Entwicklung und ihre Rolle am Arbeitsplatz sind Psychometrie und Persönlichkeitstest faszinierende Forschungsgebiete, deren Ursprünge in der klinischen Psychologie und der Personalauswahl liegen. Zunehmend werden diese Tests auch in Personalentwicklungsbereichen eingesetzt, wie z.B. im Coaching, im Bildungsmanagement, in organisatorischen Veränderungsprozessen und in der Entwicklung von Führungskräften.

Schätzungsweise vier von zehn Tests werden in der Team- und Führungskräfteentwicklung eingesetzt, wo sie oft als Instrumente zur Standortbestimmung und Selbstreflexion dienen. Hier ist das Ziel, die persönliche Entwicklung zu fördern und die Teamleistung zu verbessern. Die Marktgröße ist dabei beträchtlich: Etwa 2.500 verschiedene Tests werden jährlich einige Millionen Mal durchgeführt, was einen geschätzten Umsatz von 500 Millionen bis 2 Milliarden US-Dollar pro Jahr ergibt. Da neue Technologien die schnelle Erstellung und Verbreitung von Tests unterstützen, ist davon auszugehen, dass der Markt für Persönlichkeitstests weiter wächst. Dies bedeutet, dass MitarbeiterInnen mit großer Wahrscheinlichkeit irgendwann in ihrem beruflichen Laufbahn einmal mit Persönlichkeitstests in Berührung kommen.

Während Persönlichkeitstests jedoch oft als Wegbereiter der persönlichen Entwicklung und des organisatorischen Funktionierens angepriesen werden, können sie auch den gegenteiligen Effekt von Schubladendenken und Stereotypisierung bewirken. Eine Reihe von Persönlichkeitstests werden kritisiert, weil sie niedrige Gültigkeitswerte sowie methodische Probleme bei der Beantwortung von Selektivfragen aufweisen. Des Weiteren ist das Persönlichkeitskonstrukt selbst problematisch, da es statisch ist und wenig Raum für Veränderung und Entwicklung lässt. Bei der Personalarbeit geht es jedoch im Kern darum, nach Möglichkeiten für Lernen und persönlicher Entwicklung zu suchen. Es stellt sich also die Frage, wie ein Test, der misst „was ist“ effektiv in der Personalent-

wicklung eingesetzt werden kann, die danach strebt, Entwicklungspotenzial aufzuzeigen.

Der Einsatz von Persönlichkeitstests bleibt ein kontroverses Forschungsgebiet, in dem es wenig Literatur über deren Einsatz in der Personalentwicklung gibt. Ohne ein klares Verständnis von Persönlichkeitstests in anderen Kontexten als der Personalauswahl wirft die aktuelle Praxis ethische Bedenken hinsichtlich der Rechte und Pflichten der TestteilnehmerInnen auf. So könnten beispielsweise TestteilnehmernInnen den Eindruck bekommen, dass sie bewertet werden, auch wenn dies in einem PE-Umfeld nicht das Ziel ist. Außerdem könnten sich die TestteilnehmerInnen unwohl fühlen, wenn sie Testfeedback bei der Arbeit mit ihren Vorgesetzten und Kollegen teilen.

Diese Doktorarbeit beschäftigt sich damit, die Wahrnehmung und den Zweck von Persönlichkeitstests in der Personalentwicklung zu untersuchen, um auf eine ethische und effektive Umsetzung in diesem Bereich aufzubauen.

Die der Dissertation zugrunde liegenden Studien wurden in Westeuropa durchgeführt, mit Länderschwerpunkten in Großbritannien, Deutschland und den Niederlanden. Mit Ausnahme einiger weniger internationaler PE-PraktikerInnen aus Dänemark und den USA waren alle StudienteilnehmerInnen in einem dieser westeuropäischen Länder ansässig. Es wurde eine Reihe von qualitativen Methoden eingesetzt, die Fallstudien, qualitative Interviews und teilnehmende Beobachtungen kombinieren. Die Ergebnisse wurden mittels offenem Kodieren („open coding“), induktiver Inhaltsanalyse und qualitativer Vergleichsanalyse ausgewertet. Bei der Auswertung wurde der Rolle der kritischen Reflexion als Indikator für eine bewusste berufliche Praxis im Hinblick auf den Zweck von PE-PraktikerInnen und TestteilnehmerInnen besondere Aufmerksamkeit geschenkt.

Diese Arbeit ist in vier Studien aufgeteilt, die jeweils unterschiedliche Forschungsziele haben.

- *Studie 1* beschäftigt sich mit der Persönlichkeitstestindustrie. Diese Studie untersucht die Rolle interner und externer Interessensvertretern, den Wert psychometrischer und praktischer Überlegungen bei der Testauswahl und

den Zweck der Anwendung von Persönlichkeitstests in der betrieblichen Weiterbildung.

- Im Mittelpunkt von *Studie 2*, die eine Literaturrecherche darstellt, steht die kritische Reflexion. Die Studie untersucht die bestehende Reflexionsforschung in der Erwachsenenbildung, und wie ihre Operationalisierung die aktuelle und zukünftige Forschung beeinflusst.
- *Studie 3* konzentriert sich auf PE-Fachleute und deren Praxis beim Einsatz von Persönlichkeitstests. Diese Studie veranschaulicht individuelle Meinungsbildung (“individual meaning-making”) und organisatorische Bedeutungsbestimmung (“organizational sensemaking”) von PraktikerInnen.
- *Studie 4* widmet sich MitarbeiterInnen, die an Workshops zur Führungskräfteentwicklung teilnehmen bei denen ein Persönlichkeitstest durchgeführt wird. Diese Studie untersucht Unterschiede in den Reaktionen dieser TestteilnehmerInnen und Faktoren, die diese beeinflussen.

Studie 1: Wie lässt sich die Persönlichkeitstestindustrie mit ihren Instrumenten, Interessensvertretern und Dynamiken beschreiben?

Das Ziel der ersten Studie ist es, die Industriedynamiken rund um Persönlichkeitstests hervorzuheben, um die Rolle von Berufsverbänden zu klären und Vorschläge zur Verbesserung zu machen. Diese Mehrfachfallstudie zeigt auf, dass die Spannungen zwischen psychologischen Verbänden und Nicht-PsychologInnen ungelöst bleiben. Infolgedessen befinden sich PE-PraktikerInnen in einem Vakuum. Es gibt keine Berufsvereinigung, die für alltägliche Fragen der Anwendung oder zur Diskussion der Ethik dieser Tests in entwicklungsbezogenen Kontexten zur Beratung hinzugezogen werden kann. Psychologische Vereinigungen haben Mühe, die Durchsetzung von Teststandards unter Mitgliedern zu etablieren. Testverlage haben diese Marktlücke erkannt und bieten kostenlose Web-Veranstaltungen und Coaching-Sitzungen für diejenigen an, die mit ihrem Verfahren akkreditiert sind. Dies hat zu einer noch stärkeren Fragmentierung der Branche geführt. Es mangelt der Persönlichkeitstestindustrie an einer umfassenden Standardisierung und Kontrolle über die Testnutzung.

PE-Fachleute befinden sich in einem Feld ungelöster Branchenspannungen, in dem nicht klar ist, wie eine optimale Vorgehensweise aussehen könnte. So haben diese mit der „Umetikettierung“ ihrer Testpraxis begonnen. Anstatt den Persönlichkeitstest als „Test“ einzuführen, sprechen sie von Instrumenten zur Standortbestimmung und persönlichen Potenzialanalyse. Das Umetikettieren wird weiterhin eingesetzt, um den Test zu entmystifizieren oder herunterzuspielen und ihn stärker an den Prinzipien der Personalentwicklung und der Erwachsenenbildung auszurichten; hier werden formative Aspekte und Entwicklungspotenzial über summative Beurteilungen gestellt die den Ist-Zustand feststellen. Dabei ist zu bedenken, dass der Einsatz von Tests mit Kosten verbunden ist, die sich aus Testlizenzen und Akkreditierungsgebühren zusammensetzen. Daher ist es schwer nachvollziehbar, warum Gelder für ein Instrument ausgegeben werden, das dann als etwas anderes bezeichnet wird, nur, damit es für den Einsatz in der Personalentwicklung passend gemacht wird.

Studie 1 kommt zu dem Schluss, dass PE-Fachleute, die Persönlichkeitstests verwenden, ihr Bewusstsein für die Dynamik und Herausforderungen in diesem Bereich schärfen müssen, um mitarbeiterorientierte Ergebnisse zu erzielen. *Studie 1* empfiehlt außerdem, dass PE-Verbände den Trend, bei dem kommerzielle Verlage PraktikerInnen kostenloses Coaching anbieten, kritisch hinterfragen. PE-Verbände sollten zudem überdenken, welche Rolle sie selbst in der Persönlichkeitstestindustrie spielen wollen, da sie guten Zugang zu PraktikerInnen haben.

Studie 2: Wie hat die bestehende Literatur die kritische Reflexion konzipiert und operationalisiert?

Der Beitrag dieser Literaturrecherche besteht darin, eine Methodik zu entwickeln, die auf empirischer kritischer Reflexionsforschung basiert, um praktische Ansätze zum Erforschen der Reflexion am Arbeitsplatz zu ermitteln. Die Konzepte der inhaltlichen-, prozess- und wertenden-Reflexion („content, process and premise reflection“) wurden oft zitiert, und die transformative Lerntheorie von Mezirow (1978, 1981, 1990, 1991, 1998) und ihre Komponenten wurden von ErwachsenenbildnerInnen und PE-ForscherInnen weiterent-

wickelt¹. Mittels Literaturrecherche von 12 empirischen Studien zur kritischen Reflexion im Kontext der Erwachsenenbildung stellt diese *Studie 2* fest, dass es wenig Übereinstimmung darüber gibt, wie Reflexion operationalisiert werden kann. Bestehende Studien unterscheiden sich erheblich in Bezug auf ihre Ziele, Forschungsdesigns, Datenerhebungsmethoden und Rahmenbedingungen, in denen sie durchgeführt werden. So zielte beispielsweise eine Studie darauf ab, ein neues Kodierungsschema zu entwickeln, während eine andere Studie ein bestehendes Schema evaluierte, das in einem anderen Kontext angewendet wurde. Vor allem in der medizinischen Ausbildung lag der Schwerpunkt auf der Bewertung der kritischen Reflexionsfähigkeiten von Studierenden, während der Aspekt der Bewertung in wissenschaftlichen Arbeiten der Führungskräfteentwicklung weniger wichtig war.

Studie 2 schlägt vier mögliche Verbesserungen beim Operationalisieren der Reflexion in empirischen PE-Studien vor, nämlich die Integration verschiedener kritischer Reflexionstraditionen, die Nutzung mehrerer Datenerfassungspfade, die thematische Einbettung und die Berücksichtigung von Emotionen. Diese vier Verbesserungen wurden als Leitprinzipien bei der Durchführung der beiden verbleibenden empirischen *Studien 3 und 4* in Teilen angewandt.

Studie 3: Wie verstehen PE-Fachleute ihre eigene Persönlichkeitstestpraxis?

Diese qualitative Studie untersucht, wie erfahrene Coaches, Business-Trainer und Personalmanager das Persönlichkeitstesten in ihrer eigenen Praxis angehen, mit dem Ziel, Richtlinien für PraktikerInnen zu entwickeln, die erstmals mit Persönlichkeitstests arbeiten. Die Studie fasst die persönlichen Sichtweisen und Reflexionen von 18 Fachleuten zusammen, die im Rahmen dieser Analyse befragt wurden. Die Ergebnisse zeigen, dass PersonalentwicklerInnen sechs Strategien anwenden, wenn sie Persönlichkeitstests in ihrer Praxis einsetzen:

¹ Jack Mezirow (1923-2014), bekannt für seine Transformative Lerntheorie, war ein amerikanischer Soziologe und emeritierter Professor für Erwachsenenbildung und Weiterbildung am Teachers College der Columbia University.

- | | |
|-------------------------------------|--|
| <i>1. Ethisch-beschützend</i> | <i>4. Vorsichtig-begrüßend</i> |
| <i>2. Wissenschaftlich-selektiv</i> | <i>5. Benutzerfreundlich-pragmatisch</i> |
| <i>3. Vorsichtig-vermeidend</i> | <i>6. Fachkundig-zuvorkommend</i> |

Hier reicht das Spektrum von eher beschützend und selektiven Strategien bis hin zu Ansätzen, die als pragmatisch bezeichnet werden können, während einige PraktikerInnen Persönlichkeitstests aufgrund ethischer Bedenken ganz vermeiden.

Die Theorie der kognitiven Dissonanz² hilft teilweise zu erklären, wie PraktikerInnen von einer Strategie zur anderen wechseln, abhängig von ihrer Wahrnehmung und ihrem kritischen Bewusstsein für Persönlichkeitstests und deren Anwendung in entwicklungsbezogenen Kontexten. Dieses kritische Bewusstsein wird durch organisatorische Bedeutungsbestimmung („organizational sense-making“) im Kontext der Organisationen moduliert, mit denen diese Fachleute interagiert haben. Die Studie zeigt weiterhin, dass PE-Experten in der Regel nicht bei Null anfangen, sondern mit den bevorzugten Tests konfrontiert werden, die bereits im Unternehmen vorhanden sind oder vom Kunden gewünscht werden. Auch hier scheinen Kontext und das Spannungsfeld von Instrumenten, Interessensvertretern und Industriedynamiken, mit denen ein PE-Profi zu tun hat, die Strategie der Fachleute zu beeinflussen und möglicherweise zu bestimmen.

Nach Abschluss von *Studie 3* bleiben noch einige Fragen offen. So ist beispielsweise ungeklärt, ob die Akkreditierung von Persönlichkeitstests von Arbeitgebern und Kunden als Qualifikation dieser PersonalentwicklerInnen angesehen wird und damit auch ihren Marktwert erhöht. Darüber hinaus ist eine weitere Untersuchung erforderlich, um festzustellen, ob sich PersonalberaterInnen für spezifische Persönlichkeitstests als Alleinstellungsmerkmal entscheiden. Man könnte vermuten, dass jene Beratungsunternehmen, die nur eigenschaftsbasierte („trait-based“) Tests verwenden, als seriöser, wissenschaftlicher und professioneller wahrgenommen werden im Vergleich zu Typen-Tests. Schließlich

² Für eine Erläuterung dieser Theorie siehe Festinger, L. (2012). Theorie der Kognitiven Dissonanz (unveränderter Nachdruck der Ausgabe von 1978) Bern: Verlag Hans Huber.

bleibt abzuwarten, ob Tests in verschiedenen Kulturen unterschiedlich wahrgenommen werden.

Studie 4: Wie reagieren die TestteilnehmerInnen auf Persönlichkeitstests in Entwicklungskontexten?

Die abschließende empirische Studie zielt darauf ab, ein Modell der Reaktionen von TeilnehmerInnen in der Personalentwicklung zu konzipieren, um die Fachliteratur in entwicklungsbezogenen Kontexten voranzubringen. Hier wurden unter Verwendung von Einflussfaktoren, die aus Studien der Personalauswahl abgeleitet wurden, die Reaktionen verschiedener TestteilnehmerInnen in einem Längsschnittansatz mit elf ProbandInnen aus zwei verschiedenen Managemententwicklungsprogrammen verglichen.

Die Studie stellt fest, dass TestteilnehmerInnen den Einsatz von Persönlichkeitstests in der Führungskräfteentwicklung kritischer reflektieren als bisher in Studien zu Personalauswahl angenommen. Die Ergebnisse zeigen weiterhin, dass die Interaktion zwischen TestteilnehmerInnen und PE-Fachleuten die Reaktion der TestteilnehmerInnen in größerem Maße beeinflusst als das Instrument selbst: Durch die Auseinandersetzung mit dem Zweck der Testnutzung (*“clarity of purpose”*), der wahrgenommenen Nützlichkeit (*“perceived usefulness”*) und der respektvollen Behandlung der TestteilnehmerInnen (*“perceived respectfulness”*) steigt die Chance auf positive Teilnehmerreaktionen. Darüber hinaus beeinflussen das Bewusstsein für Annahmen (*“becoming aware of assumptions”*) und die wahrgenommene emotionale Sicherheit (*“perceived emotional safety”*), die während des Workshops erfahren wird, die Reaktionen der TestteilnehmerInnen in der Führungskräfteentwicklung. Studie 4 stellt fest, dass TestteilnehmerInnen, die sich aktiv an der Infragestellung ihrer Annahmen beteiligen und sich in der Umgebung emotional sicher fühlen, eher positiv über die Verwendung von Persönlichkeitstests urteilen, unabhängig davon, welcher spezieller Test verwendet wurde.

PersonalentwicklerInnen können alle oben genannten fünf Faktoren direkt durch ihr Handeln und ihr Verhalten während der Weiterbildung beeinflussen. Daher müssen Unternehmen, die Persönlichkeitstests in entwicklungsbezogenen Kontexten zulassen, sicherstellen, dass die eingesetzten TrainerInnen

und Coaches sich ihrer Wirkung auf die Reaktionen der TestteilnehmerInnen bewusst sind. Was als Lerninstrument für Wachstum und Fortschritt gedacht ist, kann leicht als ein Mechanismus des „Schubladendenkens“ wahrgenommen werden. Solange Berufsverbände ihre Präsenz und ihr Engagement bei PE-Fachleute nicht ausbauen, müssen Arbeitgeber in die Verantwortung gezogen werden. So sollten Test-ausführende Unternehmen Vorgaben machen, wie Persönlichkeitstests in Schulungen ethisch einwandfrei eingesetzt werden können. Diese Erkenntnis knüpft an das Fazit aus *Studie 1* an, die PE-Verbände auffordert, ihre eigene Rolle zu überdenken und zu stärken.

Gesamtbeiträge

Diese Dissertation beleuchtet verschiedene Wahrnehmungen von Persönlichkeitstests bei der Anwendung in Entwicklungskontexten. Durch die Untersuchung von Instrumenten, Interessensvertretern und Dynamiken durchleuchtete diese Dissertation die Praxis und Wahrnehmung von Persönlichkeitstests in der Personalentwicklung mit dem Ziel, auf eine ethische und effektive Umsetzung in diesem Bereich aufzubauen.

Ein Beitrag dieser Forschung ist die Förderung der kritischen Reflexion und des verstärkten Diskurses über die Ethik der Testverwendung für eine verbesserte berufliche Praxis in der Personalentwicklung. Dieser Aspekt steht in engem Zusammenhang mit dem Verständnis von Professionalität unter den PE-PraktikerInnen, die in ihrer Praxis ein gewisses Maß an Autonomie erhalten. Testverlage spielen eine wichtige Rolle in der Persönlichkeitstestindustrie, sei es durch ihre Akkreditierungsprogramme oder durch ihre beratenden Verkaufspraktiken. Dies stellt einen Interessenkonflikt dar, da Testverlage ein kommerzielles Interesse haben. Wie in *Studie 3* beschrieben, sind die PraktikerInnen in der ethischen Frage, was gute/schlechte Praxis ist, weitgehend gespalten. Dies zu hinterfragen ist ein erster Schritt, der jedoch nicht notwendigerweise dazu führt, dass PersonalentwicklerInnen ihre Persönlichkeitstestpraxis in Frage stellen. Kritische Reflexion gepaart mit einem Diskurs zwischen WissenschaftlerInnen und PraktikerInnen könnte daher von Vorteil sein, um die Bemühungen um die Gestaltung der ethischen Praxis und eines gemeinsamen Verhaltenskodex neu auszurichten.

Diese Forschung hat bereits eine Reihe informeller Diskussionen ausgelöst, die während der Forschungsvorbereitung, in Interviews, durch Workshops, in Konferenzbeiträgen und während Forschungsgruppentreffen stattfanden. Die Ergebnisse dieser Studie, insbesondere mit Bezug auf Anwendungsstrategien, sind jedoch noch nicht so dargestellt, dass sie für einen breiteren Diskurs zum einwandfreien Einsatz von Persönlichkeitstests in der Personalentwicklung genutzt werden können. Die Herausforderung für die Zukunft besteht darin, die Ergebnisse der Dissertation in didaktische Inhalte zu übersetzen, die in der Hochschul- und Managementausbildung verwendet werden können, um PE-Fachleute zu helfen, kritisch reflektierende Entscheidungen in ihrer Persönlichkeitstestpraxis zu treffen.

Schubladendenken oder Lerninstrument?

Persönlichkeitstests können als Werkzeuge zur Stereotypisierung, zum Schubladendenken und zur Abgrenzung eingesetzt werden. Als solches Instrument hebt das Testfeedback Unterschiede hervor, die als Rechtfertigung für ungeeignetes Verhalten am Arbeitsplatz genutzt werden können. Dies gilt insbesondere dann, wenn Persönlichkeitstests als Instrumente der absoluten Wahrheit eingeführt werden – ein Evangelium, das wenig Raum lässt, Annahmen in Frage zu stellen. Gleichzeitig können aber Persönlichkeitstests auch als Werkzeuge zum Lernen, Hinterfragen und zur persönlichen Entwicklung eingesetzt werden. Bei der Einführung und Behandlung auf diese Weise betont das Testfeedback Gemeinsamkeiten, die ein inklusives Arbeitsumfeld möglich machen. Dies gilt insbesondere dann, wenn zwischen den Begriffen der allgemeinen Persönlichkeit und des spezifischen Verhaltens unterschieden wird, wenn beide Begriffe entkoppelt werden und wenn versucht wird zu ermitteln, wo die Chancen für ein verbessertes Klima am Arbeitsplatz liegen. Als Lernwerkzeug kann das Wachstum und die Entwicklung einzelner Lernender durch Persönlichkeitstests unterstützt werden.

PE-Fachleute als ModeratorInnen von Persönlichkeitstests spielen eine wichtige Rolle in der Debatte über Schubladendenken-versus-Lerninstrument. Hier kommt es auf den Ansatz und der Strategie an, mit welcher der Test und sein Feedback eingeführt, verwaltet und implementiert werden. Ein Praktiker in *Studie 1* vergleicht diese Erkenntnis mit dem Besuch beim Zahnarzt: Der Patient

interessiere sich für die Qualität der Zahnarztpraxis und nicht für die Qualität der medizinischen Instrumente, die der Zahnarzt benutze. In Anlehnung an diese Analogie ist eine Zweiteilung zwischen Schubladendenken und Lerninstrument problematisch, da es vielleicht nicht die Qualität des Tests, sondern die Qualität des PE-Fachleute ist, auf die wir achten sollten.

Die Dichotomie ist auch problematisch, da sich der Arbeitszustand von einer stabileren Beschäftigung, bei der Unternehmen ArbeitnehmerInnen auf Lebenszeit einstellen, zu einem Arbeitsumfeld verlagert hat, in dem Zeitarbeit und freiberufliche Arbeit zur Norm geworden sind. Die Reaktionen auf Persönlichkeitstests sind verankert in ihrem Arbeitsumfeld, ihren Machtstrukturen und ihrem Organisationsklima. In diesem Zusammenhang ist es verständlich, dass einige TestteilnehmerInnen kritisch gegenüber Persönlichkeitstests und ihrer Anwendung in Entwicklungskontexten stehen.

References

References

- Ackerschott, H. (2000). Standards in testing and assessment: DIN 33430. In *Second International Congress on Licensure, Certification and Credentialing of Psychologists*. Oslo, Norway.
- Ackerschott, H., Gantner, N. S., & Schmitt, G. (2016). *Eignungsdiagnostik: Qualifizierte Personalentscheidungen nach DIN 33430 Mit Checklisten, Planungshilfen, Anwendungsbeispielen*. Berlin, Germany: Beuth Verlag GmbH.
- Albert, M. (1993). *Capitalism vs. capitalism*. New York, NY: John Wiley & Sons Incorporated.
- Allen, S. J., & Hartman, N. S. (2008). Leadership development: An exploration of sources of learning. *SAM Advanced Management Journal*, 73(1), 10.
- 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. <https://doi.org/10.1111/j.1744-6570.1997.tb00911.x>
- Allport, G. W. (1961). *Pattern and growth in personality*. New York, NY: Holt, Rinehart & Winston.
- Anderson, V. (2017). Criteria for evaluating qualitative research. *Human Resource Development Quarterly*, 28(2), 125–133. <https://doi.org/10.1002/hrdq.21282>
- Antonacopoulou, E. P., & Sheaffer, Z. (2014). Learning in crisis: Rethinking the relationship between Organizational Learning and Crisis Management. *Journal of Management Inquiry*, 23(1), 5–21. <https://doi.org/10.1177/1056492612472730>
- Argyris, C., & Schön, D. (1974). *Theory into practice: Increasing professional effectiveness*. San Francisco, CA: Jossey-Bass.
- Argyris, C., & Schön, D. (1996). *Organizational learning II: Theory, method and practice*. Reading, MA: Addison-Wesley.
- ATD. (2017). *State of the Industry Report* (p. 51). Alexandria, VA: The Association for Talent Development. Retrieved from <https://www.td.org/research-reports/2017-state-of-the-industry>
- Atwater, L. E., & Brett, J. F. (2005). Antecedents and consequences of reactions to developmental 360 feedback. *Journal of Vocational Behavior*, 66(3), 532–548. <https://doi.org/10.1016/j.jvb.2004.05.003>
- Atwater, L. E., Brett, J. F., & Charles, A. C. (2007). Multisource feedback: Lessons learned and implications for practice. *Human Resource Management*, 46(2), 285–307. <https://doi.org/10.1002/hrm.20161>
- Badham, R., Garrety, K., Morrigan, V., Zanko, M., & Dawson, P. (2003). Designer deviance: Enterprise and deviance in culture change programmes. *Organization*, 10(4), 707–730. <https://doi.org/10.1177/13505084030104001>

- Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology, 44*(1), 1–26. <https://doi.org/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. https://doi.org/10.1207/s15327043hup1804_3
- Barrick, M. R., Mount, M. K., & Judge, T. A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *International Journal of Selection and Assessment, 9*(1–2), 9–30. <https://doi.org/10.1111/1468-2389.00160>
- Bartram, D. (2001). Guidelines for test users: A review of national and international initiatives. *European Journal of Psychological Assessment, 17*(3), 173. <https://doi.org/10.1027//1015-5759.17.3.173>
- Bates, R., Cannonier, N., & Hatala, J.-P. (2014). Linking motivation to workplace learning transfer. In N. F. Chalofsky, T. S. Rocco, & M. L. Morris (Eds.), *Handbook of Human Resource Development* (pp. 386–401). Hoboken, NJ: John Wiley & Sons.
- Beermann, D., Kersting, M., Stegt, S., & Zimmerhofer, A. (2013). Vorurteile und Urteile zur Akzeptanz von Persönlichkeitsfragebogen. *PersonalQuarterly, 65*, 41–45.
- Behan, C. (2007). Context, creativity and critical reflection: Education in correctional institutions. *Journal of Correctional Education, 58*(2), 157–169.
- Bell, A., Kelton, J., McDonagh, N., Mladenovic, R., & Morrison, K. (2011). A critical evaluation of the usefulness of a coding scheme to categorise levels of reflective thinking. *Assessment & Evaluation in Higher Education, 36*(7), 797–815. <https://doi.org/10.1080/02602938.2010.488795>
- Benit, N., & Soellner, R. (2013). Scientist-practitioner gap in Deutschland: Eine empirische Studie am Beispiel psychologischer Testverfahren. *Zeitschrift Für Arbeits- Und Organisationspsychologie A&O, 145*–153.
- Berg-Schlosser, D., & De Meur, G. (2009). Comparative research design: Case and variable selection. In B. Rihoux & C. C. Ragin (Eds.), *Configurational comparative methods. Qualitative comparative analysis (QCA) and related techniques* (pp. 19–32). Thousand Oaks, CA: Sage Publications Ltd.
- Berings, M., Poell, R., & Gelissen, J. (2008). On-the-job learning in the nursing profession: Developing and validating a classification of learning activities and learning themes. *Personnel Review, 37*(4), 442–459. <https://doi.org/10.1108/00483480810877606>
- Berman, M. G., Jonides, J., & Nee, D. E. (2006). Studying mind and brain with fMRI. *Social Cognitive and Affective Neuroscience, 1*(2), 158–161. <https://doi.org/10.1093/scan/nsl019>
- Besharov, M. L., & Smith, W. K. (2014). Multiple institutional logics in organizations: Explaining their varied nature and implications. *Academy of Management Review, 39*(3), 364–381. <https://doi.org/10.5465/amr.2011.0431>

-
- Bierema, L. L. (1999). A model of executive women's learning and development. *Adult Education Quarterly*, 49(2), 107–121. <https://doi.org/10.1177/074171369904900203>
- Blatter, J., & Haverland, M. (2012). *Designing case studies: Explanatory approaches in small-N research*. Basingstoke, UK: Palgrave Macmillan.
- Bloom, B. S. (1956). *Taxonomy of educational objectives* (Vol. 1). New York, NY: David McKay.
- 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. <https://doi.org/www.doi.org/10.1177/0149206309352880>
- Boud, D., Keogh, R., & Walker, D. (1985). *Reflection: Turning experience into learning*. London, UK: Kogan Page.
- Bourke, J., Garr, S., van Berkel, A., & Wong, J. (2017). *Diversity and inclusion. The reality gap* (2017 Deloitte Global Human Capital Trends) (pp. 105–116). New York, NY: Deloitte University Press.
- Boyle, G. J. (1995). Myers-Briggs Type Indicator (MBTI): Some psychometric limitations. *Australian Psychologist*, 30(1), 71–74. <https://doi.org/10.1111/j.1742-9544.1995.tb01750.x>
- Briner, R. B., Denyer, D., & Rousseau, D. M. (2009). Evidence-based management: Concept cleanup time? *The Academy of Management Perspectives*, 23(4), 19–32. <https://doi.org/10.5465/amp.23.4.19>
- Briner, R. B., & Rousseau, D. M. (2011). Evidence-based I–O psychology: Not there yet. *Industrial and Organizational Psychology*, 4(1), 3–22. <https://doi.org/10.1111/j.1754-9434.2010.01287.x>
- Brookfield, S. D. (2000). The concept of critically reflective practice. In A. L. Wilson & E. R. Hayes (Eds.), *Handbook of adult and continuing education* (pp. 33–49). San Francisco, CA: Jossey-Bass.
- Broughton, J. (1977). Beyond formal operations: Theoretical thought in adolescence. *The Teachers College Record*, 79(1), 87–97.
- Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, 18(1), 32–42. <https://doi.org/10.3102/0013189X018001032>
- Bryman, A. (2008). *Social research methods* (3rd ed.). Oxford, UK: Oxford University Press.
- Burke, L. A., & Hutchins, H. M. (2007). Training transfer: An integrative literature review. *Human Resource Development Review*, 6(3), 263–296. <https://doi.org/10.1177/1534484307303035>
- Cacioppe, R. (1998). An integrated model and approach for the design of effective leadership development programs. *Leadership & Organization Development Journal*, 19(1), 44–53. <https://doi.org/10.1108/01437739810368820>
- Callahan, J. L. (2004). Breaking the cult of rationality: Mindful awareness of emotion in the critical theory classroom. *New Directions for Adult and Continuing Education*, 2004(102), 75–83. <https://doi.org/10.1002/ace.140>

- Callahan, J. L. (2010). Constructing a manuscript: Distinguishing integrative literature reviews and conceptual and theory articles. *Human Resource Development Review, 9*(3), 300–304. <https://doi.org/10.1177/1534484310371492>
- Callahan, J. L., & McCollum, E. E. (2002). Conceptualizations of emotion research in organizational contexts. *Advances in Developing Human Resources, 4*(1), 4–21. <https://doi.org/10.1177/1523422302004001002>
- Caramani, D. (2008). *Introduction to the comparative method with Boolean algebra*. Thousand Oaks, CA: Sage Publications Ltd.
- Carducci, G., Rizzo, G., Monti, D., Palumbo, E., & Morisio, M. (2018). TwitPersonality: Computing personality traits from tweets using word embeddings and supervised learning. *Information, 9*(5), 127. <https://doi.org/10.3390/info9050127>
- Carlner, S., & Hamlin, B. (2015). Certification of HRD professionals. In R. F. Poell, T. S. Rocco, & G. L. Roth (Eds.), *The Routledge companion to Human Resource Development* (pp. 343–366). Abingdon, UK: Routledge.
- Chalofsky, N. F. (2007). The seminal foundation of the discipline of HRD: People, learning, and organizations. *Human Resource Development Quarterly, 18*(3), 431–442. <https://doi.org/10.1002/hrdq.1212>
- Chamorro-Premuzic, T., & Furnham, A. (2009). Mainly openness: The relationship between the Big Five personality traits and learning approaches. *Learning and Individual Differences, 19*(4), 524–529. <https://doi.org/10.1016/j.lindif.2009.06.004>
- Chen, A., & Potenza, A. (2018, March 20). Cambridge Analytica's Facebook data abuse shouldn't get credit for Trump. *The Verge, 1*–10.
- Chirema, K. D. (2007). The use of reflective journals in the promotion of reflection and learning in post-registration nursing students. *Nurse Education Today, 27*(3), 192–202. <https://doi.org/10.1016/j.nedt.2006.04.007>
- Choi, M., & Roulston, K. (2015). Learning transfer in practice: A qualitative study of medical professionals' perspectives. *Human Resource Development Quarterly, 26*(3), 249–273. <https://doi.org/10.1002/hrdq.21209>
- Church, A. H., Gilbert, M., Oliver, D. H., Paquet, K., & Surface, C. (2002). The role of technology in organization development and change. *Advances in Developing Human Resources, 4*(4), 493–511. <https://doi.org/10.1177/152342202237525>
- Clarke, A. (1995). Professional development in practicum settings: Reflective practice under scrutiny. *Teaching and Teacher Education, 11*(3), 243–261. [https://doi.org/10.1016/0742-051X\(94\)00028-5](https://doi.org/10.1016/0742-051X(94)00028-5)
- Clinebell, S., & Stecher, M. (2003). Teaching teams to be teams: An exercise using the Myers-Briggs® Type Indicator and the five-factor personality traits. *Journal of Management Education, 27*(3), 362–383. <https://doi.org/10.1177/1052562903027003006>

-
- Converse, P. D., Oswald, F. L., Imus, A., Hedricks, C., Roy, R., & Butera, H. (2008). Comparing personality test formats and warnings: Effects on criterion-related validity and test-taker reactions. *International Journal of Selection and Assessment, 16*(2), 155–169. <https://doi.org/10.1111/j.1468-2389.2008.00420.x>
- Cooper, S. (2014). Transformative evaluation: Organisational learning through participative practice. *The Learning Organization, 21*(2), 146–157. <https://doi.org/10.1108/TLO-03-2013-0003>
- Cope, J. (2003). Entrepreneurial learning and critical reflection discontinuous events as triggers for 'higher-level' learning. *Management Learning, 34*(4), 429–450. <https://doi.org/10.1177/1350507603039067>
- Cope, J., & Watts, G. (2000). Learning by doing: An exploration of experience, critical incidents and reflection in entrepreneurial learning. *International Journal of Entrepreneurial Behaviour & Research, 6*(3), 104–124. <https://doi.org/10.1108/13552550010346208>
- Cranton, P. (2006a). Fostering authentic relationships in the transformative classroom. *New Directions for Adult and Continuing Education, 2006*(109), 5–13. <https://doi.org/10.1002/ace.203>
- Cranton, P. (2006b). *Understanding and promoting Transformative Learning: A guide for educators of adults*. San Francisco, CA: Jossey-Bass.
- Cranton, P., & Carusetta, E. (2004). Developing authenticity as a transformative process. *Journal of Transformative Education, 2*(4), 276–293. <https://doi.org/10.1177/1541344604267898>
- Cseh, M. (2003). Facilitating learning in multicultural teams. *Advances in Developing Human Resources, 5*(1), 26–40. <https://doi.org/10.1177/1523422302239181>
- Deal, J. J., Altman, D. G., & Rogelberg, S. G. (2010). Millennials at work: What we know and what we need to do (if anything). *Journal of Business and Psychology, 25*(2), 191–199. <https://doi.org/10.1007/s10869-010-9177-2>
- Dewey, J. (1933). *How we think: A restatement of the relations of reflective thinking to the educative process*. Lexington, MA: D.C. Heath & Co.
- Diekmann, J., & König, C. J. (2015). Personality testing in personnel selection: Love it? Leave it? Understand it! In J. K. Oostrom & I. Nikolaou (Eds.), *Employee Recruitment, Selection, and Assessment*. Hove, UK: Psychology Press.
- DiMilia, L. (2004). Australian management selection practices: Closing the gap between research findings and practice. *Asia Pacific Journal of Human Resources, 42*(2), 214–228. <https://doi.org/10.1177/1038411104045364>
- DIN. (2002). *DIN 33430: Anforderungen an Verfahren und deren Einsatz bei berufsbezogenen Eignungsbeurteilungen*. Berlin, Germany: Beuth.
- Dirkx, J. M. (2008). The meaning and role of emotions in adult learning. *New Directions for Adult and Continuing Education, 2008*(120), 7–18. <https://doi.org/10.1002/ace.311>

- Dirkx, J. M., Mezirow, J., & Cranton, P. (2006). Musings and reflections on the meaning, context, and process of Transformative Learning: A dialogue between John M. Dirkx and Jack Mezirow. *Journal of Transformative Education, 4*(2), 123–139. <https://doi.org/10.1177/1541344606287503>
- Dyment, J. E., & O’Connell, T. S. (2011). Assessing the quality of reflection in student journals: A review of the research. *Teaching in Higher Education, 16*(1), 81–97. <https://doi.org/10.1080/13562517.2010.507308>
- Eisenhardt, K. M. (1989). Building theories from case study research. *Academy of Management Review, 14*(4), 532–550. <https://doi.org/10.5465/amr.1989.4308385>
- Ellinger, A. D. (2004). The concept of self-directed learning and its implications for human resource development. *Advances in Developing Human Resources, 6*(2), 158–177. <https://doi.org/10.1177/1523422304263327>
- Ellinger, A. D., Watkins, K. E., & Marsick, V. J. (2005). Case study research methods. In R. A. Swanson & E. F. Holton (Eds.), *Research in organizations: Foundations and methods of inquiry* (pp. 327–350).
- Emre, M. (2018). *The personality brokers - The strange history of Myers-Briggs and the birth of personality testing*. New York, NY: Doubleday.
- Enckell, H. (2010). Reflection in psychoanalysis: On symbols and metaphors. *The International Journal of Psychoanalysis, 91*(5), 1093–1114. <https://doi.org/10.1111/j.1745-8315.2010.00320.x>
- Ennis, R. H. (1993). Critical thinking assessment. *Theory into Practice, 32*(3), 179–186. <https://doi.org/10.1080/00405849309543594>
- Evers, A. (1996). Regulations concerning test qualifications and test use in The Netherlands. *European Journal of Psychological Assessment, 12*(2), 153–159. <https://doi.org/10.1027/1015-5759.12.2.153>
- Evers, A., Muñiz, J., Bartram, D., Boben, D., Egeland, J., Fernández-Hermida, J. R., ... others. (2012). Testing practices in the 21st century. *European Psychologist. https://doi.org/10.1027/1016-9040/a000102*
- Felder, A. (2005, June). Drum teste Tests, mit denen du getestet wirst. *HR Today*, pp. 1–2.
- Fenwick, T. J. (2000a). Expanding conceptions of experiential learning: A review of the five contemporary perspectives on cognition. *Adult Education Quarterly, 50*(4), 243–272. <https://doi.org/10.1177/07417130022087035>
- Fenwick, T. J. (2000b). Putting meaning into workplace learning. In A. L. Wilson & E. R. Hayes (Eds.), *Handbook of Adult and Continuing Education* (pp. 294–311). San Francisco, CA: John Wiley & Sons.
- Fenwick, T. J. (2003). Reclaiming and re-embodying experiential learning through complexity science. *Studies in the Education of Adults, 35*(2), 123–141. <https://doi.org/10.1080/02660830.2003.11661478>
- Fenwick, T. J. (2014). Sociomateriality in medical practice and learning: attuning to what matters. *Medical Education, 48*(1), 44–52. <https://doi.org/10.1111/medu.12295>

-
- Fenwick, T. J. (2016). *Professional responsibility and professionalism: A sociomaterial examination*. New York, NY: Routledge.
- Fenwick, T. J., & Edwards, R. (2017). How sociomaterial approaches could support higher education as a critical practice. *Theorising Learning to Teach in Higher Education*, 34–52.
- Festinger, L. (1962). *A theory of cognitive dissonance* (Vol. 2). Palo Alto, CA: Stanford University Press.
- Fielding, N., & Thomas, H. (2001). Qualitative interviewing. In N. Gilbert (Ed.), *Researching Social Life*. London, UK: Sage Publications Ltd.
- Fisher, S. G., Hunter, T., & Macrosson, W. (2001). A validation study of Belbin's team roles. *European Journal of Work and Organizational Psychology*, 10(2), 121–144. <https://doi.org/10.1080/13594320143000591>
- Fluckinger, C. D. (2010). *Measurement of Big Five personality via Q-Sort: Comparison with a Likert measure and test-taker perceptions and reactions* (PhD thesis). University of Akron, Akron, OH.
- Fluckinger, C. D., & Snell, A. F. (2016). User reactions to personality assessment. In U. Kumar (Ed.), *The Wiley Handbook of Personality Assessment* (1st ed., pp. 361–373). Chichester, UK: John Wiley & Sons.
- Ford, J., & Harding, N. (2007). Move over management: We are all leaders now. *Management Learning*, 38(5), 475–493. <https://doi.org/10.1177/1350507607083203>
- Franz, N. (2010). Catalyzing employee change with transformative learning. *Human Resource Development Quarterly*, 21(1), 113–118. <https://doi.org/10.1002/hrdq.20033>
- Freeman, R. E., Harrison, J. S., Wicks, A. C., Parmar, B. L., & De Colle, S. (2010). *Stakeholder theory: The state of the art*. Cambridge, UK: Cambridge University Press.
- Friedman, H. S., & Schustack, M. W. (2010). *Personality: Classic theories and modern research* (5th ed.). Harlow, UK: Pearson.
- Furnham, A. (2004). Personality and organisation: European perspectives on personality assessment in organisations. *Personality and Organisations*, 25–59.
- Furnham, A. (2008). HR professionals' beliefs about, and knowledge of, assessment techniques and psychometric tests. *International Journal of Selection and Assessment*, 16(3), 300–305. <https://doi.org/10.1111/j.1468-2389.2008.00436.x>
- Furnham, A. (2017). Does it matter who we are? Personality at work (Ch 17). In N. Chmiel, F. Fraccaroli, & M. Sverke (Eds.), *An introduction to work and organizational psychology: An international perspective* (3rd ed., pp. 317–334). Hoboken, NJ: John Wiley & Sons.
- Furnham, A., & Crump, J. (2005). Personality traits, types, and disorders: an examination of the relationship between three self-report measures. *European Journal of Personality*, 19(3), 167–184. <https://doi.org/10.1002/per.543>
- Furnham, A., Eracleous, A., & Chamorro-Premuzic, T. (2009). Personality, motivation and job satisfaction: Hertzberg meets the Big Five. *Journal of Managerial Psychology*, 24(8), 765–779. <https://doi.org/10.1108/02683940910996789>

- Furnham, A., & Jackson, C. J. (2011). Practitioner reactions to work-related psychological tests. *Journal of Managerial Psychology, 26*(7), 549–565. <https://doi.org/10.1108/02683941111164472>
- Garavan, T. N., McGuire, D., & Lee, M. (2015). Reclaiming the “D” in HRD: A typology of development conceptualizations, antecedents, and outcomes. *Human Resource Development Review, 14*(4), 359–388. <https://doi.org/10.1177/1534484315607053>
- Gardner, W. L., & Martinko, M. J. (1996). Using the Myers-Briggs Type Indicator to study managers: A literature review and research agenda. *Journal of Management, 22*(1), 45–83. <https://doi.org/10.1177/014920639602200103>
- Gayle, B. M., Cortez, D., & Preiss, R. W. (2013). Safe spaces, difficult dialogues, and critical thinking. *International Journal for the Scholarship of Teaching and Learning, 7*(2), 5. <https://doi.org/10.20429/ijstl.2013.070205>
- Gegenfurtner, A., Veermans, K., Festner, D., & Gruber, H. (2009). Motivation to transfer training: An integrative literature review. *Human Resource Development Review, 8*(3), 403–423. <https://doi.org/10.1177/1534484309335970>
- Gerrig, R. J., & Zimbardo, P. G. (2009). *Psychology and Life* (19th ed.). Harlow, UK: Pearson Education.
- Geue, P. E. (2018). Positive practices in the workplace: Impact on team climate, work engagement, and task performance. *The Journal of Applied Behavioral Science, 54*(3), 272–301. <https://doi.org/10.1177/0021886318773459>
- Gilliland, S. W. (1993). The perceived fairness of selection systems: An organizational justice perspective. *Academy of Management Review, 18*(4), 694–734. <https://doi.org/10.5465/amr.1993.9402210155>
- Gilliland, S. W. (1994). Effects of procedural and distributive justice on reactions to a selection system. *Journal of Applied Psychology, 79*(5), 691. <https://doi.org/10.1037/0021-9010.79.5.691>
- Gilliland, S. W. (1995). Fairness from the applicant’s perspective: Reactions to employee selection procedures. *International Journal of Selection and Assessment, 3*(1), 11–18. <https://doi.org/10.1111/j.1468-2389.1995.tb00002.x>
- Gilliland, S. W., Benson III, L., & Schepers, D. H. (1998). A rejection threshold in justice evaluations: Effects on judgment and decision-making. *Organizational Behavior and Human Decision Processes, 76*(2), 113–131. <https://doi.org/10.1006/obhd.1998.2801>
- Gilliland, S. W., & Steiner, D. D. (2012). Applicant reactions to testing and selection. In N. Schmitt (Ed.), *The Oxford handbook of personnel assessment and selection*. Oxford University Press.
- Giroux, H. A. (2005). *Border crossings: Cultural workers and the politics of education*. New York, NY: Taylor & Francis Group.
- Glaser, B. (1998). *Doing grounded theory: Issues and discussions*. Mill Valley, CA: Sociology Press.
- Glaser, B., & Strauss, A. (1967). *The discovery of grounded theory: Strategies for qualitative research*. London, UK: Aldine.

- Goldberg, L. R. (1990). An alternative "description of personality": The Big-Five factor structure. *Journal of Personality and Social Psychology*, 59(6), 1216–1229. <https://doi.org/10.1037/0022-3514.59.6.1216>
- Goodstein, L. D., & Lanyon, R. I. (1999). Applications of personality assessment to the workplace: A review. *Journal of Business and Psychology*, 13(3), 291–322. <https://doi.org/10.1023/A:1022941331649>
- Goodstein, L. D., & Prien, E. P. (2006). *Using individual assessments in the workplace: A practical guide for HR professionals, trainers, and managers*. San Francisco, CA: John Wiley & Sons.
- Gray, D. E. (2007). Facilitating management learning: Developing critical reflection through reflective tools. *Management Learning*, 38(5), 495–517. <https://doi.org/10.1177/1350507607083204>
- Gray, E. (2015, June 11). How high is your XQ? *Time Magazine*, 40–46.
- Greeno, J. G. (1997). Theories and practices of thinking and learning to think. *American Journal of Education*, 106, 85–126. <https://doi.org/10.1086/444177>
- Gulliford, A. R. (1991). The role of personality in assessing management potential. *Personnel Review*, 20(1), 25–31. <https://doi.org/10.1108/00483489110139135>
- Habermas, J. (1971). *Toward a rational society: Student protest, science, and politics* (Vol. 404). Boston, MA: Beacon Press.
- Hagemester, C., Kersting, M., & Stemmler, G. (2012). Test reviewing in Germany. *International Journal of Testing*, 12(2), 185–194. <https://doi.org/10.1080/15305058.2012.657922>
- Hagemester, C., Lang, F., & Kersting, M. (2010). Einstellung von Psychologinnen und Psychologen in Deutschland zu Tests. *Report Psychologie*, 35, 428–439.
- Harland, L. K. (2003). Using personality tests in leadership development: Test format effects and the mitigating impact of explanations and feedback. *Human Resource Development Quarterly*, 14(3), 285–301. <https://doi.org/10.1002/hrdq.1067>
- Hausknecht, J. P., Day, D. V., & Thomas, S. C. (2004). Applicant reactions to selection procedures: An updated model and meta-analysis. *Personnel Psychology*, 57(3), 639–683. <https://doi.org/10.1111/j.1744-6570.2004.00003.x>
- Hinojosa, A. S., Gardner, W. L., Walker, H. J., Coglisier, C., & Gullifor, D. (2017). A review of cognitive dissonance theory in management research: Opportunities for further development. *Journal of Management*, 43(1), 170–199. <https://doi.org/10.1177/0149206316668236>
- Hogan, J., & Holland, B. (2003). Using theory to evaluate personality and job-performance relations: A socioanalytic perspective. *Journal of Applied Psychology*, 88(1), 100. <https://doi.org/www.doi.org/10.1037/0021-9010.88.1.100>
- Hossiep, R., Schecke, J., & Weiss, S. (2015). Zum Einsatz von persönlichkeitsorientierten Fragebogen. *Psychologische Rundschau*, 66(2). <https://doi.org/10.1026/0033-3042/a000235>

- Hough, L. M., & Oswald, F. L. (2005). They're right, well... mostly right: Research evidence and an agenda to rescue personality testing from 1960s insights. *Human Performance, 18*(4), 373–387.
- Hough, L. M., & Oswald, F. L. (2008). Personality testing and industrial–organizational psychology: Reflections, progress, and prospects. *Industrial and Organizational Psychology, 1*(3), 272–290. https://doi.org/10.1207/s15327043hup1804_4
- Howie, P., & Bagnall, R. (2015). A critical comparison of transformation and deep approach theories of learning. *International Journal of Lifelong Education, 3*(3), 1–18. <https://doi.org/10.1080/02601370.2014.1000409>
- Jarzabkowski, P., Balogun, J., & Seidl, D. (2007). Strategizing: The challenges of a practice perspective. *Human Relations, 60*(1), 5–27. <https://doi.org/10.1177/0018726707075703>
- Jordan, S. (2010). Learning to be surprised: How to foster reflective practice in a high-reliability context. *Management Learning, 41*(4), 391–413. <https://doi.org/10.1177/1350507609357388>
- Jung, C. G. (1923). *Psychological types: Or, the psychology of individuation*. London, UK: Routledge and K. Paul.
- Justice, S. B., Lundgren, H., Marsick, V. J., Poell, R. F., & Yorks, L. (2018). *Operationalizing reflection in experience-based workplace learning: A hybrid approach*. Working paper, Teachers College, Columbia University.
- Justice, S. B., & Yorks, L. (2018). Incidental learning as an enacted encounter with materiality. *New Directions for Adult and Continuing Education, 2018*(159), 91–102. <https://doi.org/10.1002/ace.20289>
- Kahnweiler, W. M. (2009). HRD as a profession: Current status and future directions. *Human Resource Development Quarterly, 20*(2), 219–229. <https://doi.org/10.1002/hrdq.20011>
- Keevers, L., & Treleaven, L. (2011). Organizing practices of reflection: A practice-based study. *Management Learning, 42*(5), 505–520. <https://doi.org/10.1177/1350507610391592>
- Keirse, D., & Bates, M. (1984). *Please understand me: Character and temperament types* (5th ed.). Del Mar, CA: Prometheus Nemesis Book Company.
- Kember, D., Jones, A., Loke, A., McKay, J., Tse, H., Sinclair, K., ... Yeung, E. (1999). Determining the level of reflective thinking from students' written journals using a coding scheme based on the work of Mezirow. *International Journal of Lifelong Education, 18*(1), 18–30. <https://doi.org/10.1080/026013799293928>
- Kember, D., Leung, D. Y., Jones, A., Loke, A. Y., McKay, J., Sinclair, K., ... Wong, M. (2000). Development of a questionnaire to measure the level of reflective thinking. *Assessment & Evaluation in Higher Education, 25*(4), 381–395. <https://doi.org/10.1080/713611442>
- Kersting, M. (2008a). *Qualität in der Diagnostik und Personalauswahl: Der DIN Ansatz*. Göttingen: Hogrefe Verlag.
- Kersting, M. (2008b). Zur Akzeptanz von Intelligenz- und Leistungstests. *Report Psychologie, 33*(9), 420–433.

-
- Kersting, M. (2010). Akzeptanz von Assessment Centern: Was kommt an und worauf kommt es an. *Wirtschaftspsychologie, 12*, 58–65.
- King, K. P. (2004). Both sides now: Examining transformative learning and professional development of educators. *Innovative Higher Education, 29*(2), 155–174. <https://doi.org/10.1023/B:IHIE.0000048796.60390.5f>
- King, P. M., & Kitchener, K. S. (1994). *Developing reflective judgment: Understanding and promoting intellectual growth and critical thinking in adolescents and adults*. San Francisco, CA: Jossey-Bass Publishers.
- Kitchener, K. S., King, P. M., & DeLuca, S. (2006). Development of reflective judgment in adulthood. In C. Hoare (Ed.), *Handbook of adult development and learning* (pp. 73–98). New York, NY: Oxford University Press.
- Kitchenham, A. (2006). Teachers and technology: A transformative journey. *Journal of Transformative Education, 4*(3), 202–225. <https://doi.org/10.1177/1541344606290947>
- Kitchenham, A. (2008). The evolution of John Mezirow's Transformative Learning Theory. *Journal of Transformative Education, 6*(2), 104–123. <https://doi.org/10.1177/1541344608322678>
- Kitchenham, A., & Chasteauneuf, C. (2009). An application of Mezirow's Critical Reflection Theory to electronic portfolios. *Journal of Transformative Education, 7*(3), 230–244. <https://doi.org/10.1177/1541344610383287>
- Kline, P. (1976). *Psychological testing: the measurement of intelligence, ability and personality*. London, UK: Malaby Press.
- Konradt, U., Garbers, Y., Böge, M., Erdogan, B., & Bauer, T. N. (2017). Antecedents and consequences of fairness perceptions in personnel selection: a 3-year longitudinal study. *Group & Organization Management, 42*(1), 113–146. <https://doi.org/10.1177/1059601115617665>
- Koole, S., Dornan, T., Aper, L., Scherpbier, A., Valcke, M., Cohen-Schotanus, J., & Derese, A. (2011). Factors confounding the assessment of reflection: A critical review. *BMC Medical Education, 11*(1), 104. <https://doi.org/10.1186/1472-6920-11-104>
- Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences, 201218772*. <https://doi.org/10.1073/pnas.1218772110>
- Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory into Practice, 41*(4), 212–218. https://doi.org/10.1207/s15430421tip4104_2
- Kreber, C. (1999). A course-based approach to the development of teaching-scholarship: a case study. *Teaching in Higher Education, 4*(3), 309–325. <https://doi.org/10.1080/1356251990040301>
- Kreber, C. (2004). An analysis of two models of reflection and their implications for educational development. *International Journal for Academic Development, 9*(1), 29–49.
- Kreber, C. (2005). Reflection on teaching and the scholarship of teaching: Focus on science instructors. *Higher Education, 50*(2), 323–359. <https://doi.org/10.1007/s10734-004-6360-2>

- Kreber, C. (2012). Critical Reflection and Transformative Learning. In E. W. Taylor & P. Cranton (Eds.), *The handbook of transformative learning: Theory, research, and practice* (pp. 323–341). San Francisco, CA: Jossey-Bass.
- Kreber, C., & Castleden, H. (2009). Reflection on teaching and epistemological structure: Reflective and critically reflective processes in 'pure/soft' and 'pure/hard' fields. *Higher Education*, 57(4), 509–531. <https://doi.org/10.1007/s10734-008-9158-9>
- Kreber, C., & Cranton, P. (2000). Exploring the Scholarship of Teaching. *Journal of Higher Education*, 71(4), 476–495. <https://doi.org/10.2307/2649149>
- Krumboltz, J. D. (2009). The happenstance learning theory. *Journal of Career Assessment*, 17(2), 135–154. <https://doi.org/10.1177/1069072708328861>
- Kucukaydin, I., & Cranton, P. (2013). Critically questioning the discourse of Transformative Learning Theory. *Adult Education Quarterly*, 63(1), 43–56. <https://doi.org/10.1177/0741713612439090>
- Kuipers, B. S., Higgs, M. J., Tolkacheva, N. V., & de Witte, M. C. (2009). The influence of Myers-Briggs type indicator profiles on team development processes: An empirical study in the manufacturing industry. *Small Group Research*. <https://doi.org/10.1177/1046496409333938>
- 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. <https://doi.org/10.1006/jvbe.2001.1854>
- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge, MA: Cambridge University Press. Retrieved from <https://doi.org/10.1017/CBO9780511815355>
- Lawrence, P. (2015). A best practice model for the effective deployment of 360 feedback. *Development and Learning in Organizations: An International Journal*, 29(6), 13–16. <https://doi.org/10.1108/DLO-02-2015-0012>
- Layne, C., & Ally, G. (1980). How and why people accept personality feedback. *Journal of Personality Assessment*, 44(5), 541–546. https://doi.org/10.1207/s15327752jpa4405_16
- Lee, M. (2001). A refusal to define HRD. *Human Resource Development International*, 4(3), 327–341. <https://doi.org/10.1080/13678860110059348>
- Lewis, M. W. (2000). Exploring paradox: Toward a more comprehensive guide. *Academy of Management Review*, 25(4), 760–776. <https://doi.org/10.2307/259204>
- Lewis, M. W., & Smith, W. K. (2014). Paradox as a metatheoretical perspective: Sharpening the focus and widening the scope. *The Journal of Applied Behavioral Science*, 50(2), 127–149. <https://doi.org/10.1177/0021886314522322>
- Li, L., Gray, D. E., Lockwood, A. J., & Buhalis, D. (2013). Learning about managing the business in the hospitality industry. *Human Resource Development Quarterly*, 24(4), 525–559. <https://doi.org/10.1002/hrdq.21171>

-
- Lievens, F., van Dam, K., & Anderson, N. (2002). Recent trends and challenges in personnel selection. *Personnel Review*, 31(5), 580–601. <https://doi.org/10.1108/00483480210438771>
- Liimatainen, L., Poskiparta, M., Karhila, P., & Sjögren, A. (2001). The development of reflective learning in the context of health counselling and health promotion during nurse education. *Journal of Advanced Nursing*, 34(5), 648–658.
- Liimatainen, L., Poskiparta, M., Karhila, P., & Sjogren, A. (2001). The development of reflective learning in the context of health counselling and health promotion during nurse education. *Journal of Advanced Nursing*, 34(5), 648–658. <https://doi.org/10.1046/j.1365-2648.2001.01794.x>
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry* (Vol. 75). Newbury Park, CA: Sage.
- London, M., & Sessa, V. I. (2006). Group Feedback for Continuous Learning. *Human Resource Development Review*, 5(3), 303–329.
- Lucas, R. E., & Donnellan, M. B. (2011). Personality development across the life span: Longitudinal analyses with a national sample from Germany. *Journal of Personality and Social Psychology*, 101(4), 847–861. <https://doi.org/10.1037/a0024298>
- Ludeman, K. (1995). Motorola's HR learns the value of teams firsthand. *Personnel Journal*, 74(6), 117–118, 120, 123.
- Lundgren, H. (2012). Schubladendenken oder Lerninstrument? Persönlichkeitstests und deren Anwendung in der betrieblichen Weiterbildung. *Magazin Erwachsenenbildung.At*, 17. Retrieved from <http://erwachsenenbildung.at/magazin/12-17/meb12-17.pdf>
- Lundgren, H. (2013). On critical reflection: Functionalizing “levels of reflection” within empirical Transformative Learning research. In *Proceedings of the 32nd annual conference (2013)* (pp. 348–355). Victoria, BC, Canada.
- Lundgren, H. (2014a). Me, myself, and I: On the role of self-reflection in adult education. Presented at the Adult Education Research Conference, Harrisburg, PA: New Prairie Press. Retrieved from <http://newprairiepress.org/aerc/2014/papers/49>
- Lundgren, H. (2014b). Testing as Reflecting? Preliminary findings from a study involving personality testing in CVET. In *Proceedings of the European Society for Research on the Education of Adults*. Berlin, Germany.
- Lundgren, H., Bang, A., Justice, S. B., Marsick, V. J., Poell, R. F., Yorks, L., ... Sung, S. (2017). Conceptualizing reflection in experience-based workplace learning. *Human Resource Development International*, 20(4), 305–326. <https://doi.org/10.1080/13678868.2017.1308717>
- Lundgren, H., Kroon, B., & Poell, R. F. (2017). Personality testing and workplace training: Exploring stakeholders, products and purpose in Western Europe. *European Journal of Training and Development*, 41(3), 198–221. <https://doi.org/10.1108/EJTD-03-2016-0015>

- Lundgren, H., & Poell, R. F. (2016). On critical reflection: A review of Mezirow's theory and its operationalization. *Human Resource Development Review, 15*(1), 3–28. <https://doi.org/10.1177/1534484315622735>
- Lundin, J., & Nuldén, U. (2007). Talking about tools: Investigating learning at work in police practice. *Journal of Workplace Learning, 19*(4), 222–239. <https://doi.org/10.1108/13665620710747915>
- Lyons, S., & Kuron, L. (2014). Generational differences in the workplace: A review of the evidence and directions for future research. *Journal of Organizational Behavior, 35*(S1), S139–S157. <https://doi.org/10.1002/job.1913>
- Maines, D. R. (1991). Reflections, framings, and appreciations. In D. R. Maines (Ed.), *Social organization and social process: essays in honor of Anselm Strauss* (pp. 3–10). New York, NY: Aldine De Gruyter.
- Maitlis, S., & Christianson, M. (2014). Sense-making in organizations: Taking stock and moving forward. *The Academy of Management Annals, 8*(1), 57–125. <https://doi.org/10.1080/19416520.2014.873177>
- Mälkki, K. (2010). Building on Mezirow's theory of Transformative Learning: Theorizing the challenges to reflection. *Journal of Transformative Education, 8*(1), 42–62. <https://doi.org/10.1177/1541344611403315>
- Mälkki, K. (2012). Rethinking disorienting dilemmas within real-life crises: The role of reflection in negotiating emotionally chaotic experiences. *Adult Education Quarterly, 62*(3), 207–229. <https://doi.org/10.1177/0741713611402047>
- Mann, K., Gordon, J., & MacLeod, A. (2009). Reflection and reflective practice in health professions education: A systematic review. *Advances in Health Sciences Education, 14*(4), 595–621. <https://doi.org/10.1007/s10459-007-9090-2>
- Marsick, V. J., & Maltbia, T. E. (2009). The transformative potential of action learning conversations. In J. Mezirow & E. W. Taylor (Eds.), *Transformative learning in practice: Insights from community, workplace, and higher education* (pp. 160–171). San Francisco, CA: Jossey-Bass.
- Marsick, V. J., Watkins, K., & O'Connor, B. (2011). Researching workplace learning in the United States. In M. Malloch, L. Cairns, K. Evans, & B. O'Connor (Eds.), *The SAGE handbook of workplace learning* (pp. 198–209). London, UK: Sage Publications Ltd.
- Massenberg, A.-C., Schulte, E.-M., & Kauffeld, S. (2017). Never too early: Learning transfer system factors affecting motivation to transfer before and after training programs. *Human Resource Development Quarterly, 28*(1), 55–85. <https://doi.org/10.1002/hrdq.21256>
- Matz, D. C., & Wood, W. (2005). Cognitive dissonance in groups: The consequences of disagreement. *Journal of Personality and Social Psychology, 88*(1), 22. <https://doi.org/10.1037/0022-3514.88.1.22>
- Maudsley, D. B. (1979). *A theory of meta-learning and principles of facilitation: An organismic perspective* (PhD thesis). University of Toronto, Toronto.
- McAdams, D. P. (1997). A conceptual history of personality psychology. In R. Hogan, J. Johnson, & S. Briggs (Eds.), *Handbook of personality psychology* (pp. 3–39). San Diego, CA: Academic Press.

-
- McCarthy, A. M., & Garavan, T. N. (1999). Developing self-awareness in the managerial career development process: The value of 360-degree feedback and the MBTI. *Journal of European Industrial Training*, 23(9), 437–445. <https://doi.org/10.1108/03090599910302613>
- McCarthy, A. M., & Garavan, T. N. (2001). 360 feedback process: Performance, improvement and employee career development. *Journal of European Industrial Training*, 25(1), 5–32. <https://doi.org/10.1108/03090590110380614>
- McCrae, R. R., & Costa, P. T. (1999). A five-factor theory of personality. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of Personality: Theory and Research* (2nd ed., pp. 139–153). New York, NY: The Guildford Press.
- McCrae, R. R., & Costa, P. T. (2003). *Personality in adulthood: A five-factor theory perspective*. New York, NY: Guilford Press.
- McCrae, R. R., & McCrae, P. T. (1996). Toward a new generation of personality theories: Theoretical contexts for the five-factor model. In J. S. Wiggins (Ed.), *The five-factor model of personality: Theoretical perspectives* (pp. 51–87). New York, NY: Guilford Publications.
- McGurk, J., & Belliveau, V. (2012). *Learning and talent development 2012* (Annual Survey Report). London, UK: Chartered Institute of Personnel and Development (CIPD).
- Mertens, D. M. (2014). *Research and evaluation in education and psychology: Integrating diversity with quantitative, qualitative, and mixed methods* (4th ed.). Los Angeles, CA: Sage Publications Ltd.
- Mezirow, J. (1978). Perspective transformation. *Adult Education Quarterly*, 28(2), 100–110. <https://doi.org/10.1177/074171367802800202>
- Mezirow, J. (1981). A critical theory of adult learning and education. *Adult Education Quarterly*, 32(1), 3–24. <https://doi.org/10.1177/074171368103200101>
- Mezirow, J. (1990). *Fostering critical reflection in adulthood: A guide to transformative and emancipatory learning*. San Francisco, CA: Jossey-Bass.
- Mezirow, J. (1991). *Transformative dimensions of adult education*. San Francisco, CA: Jossey-Bass.
- Mezirow, J. (1998). On critical reflection. *Adult Education Quarterly*, 48(3), 185–198. <https://doi.org/10.1177/074171369804800305>
- Mezirow, J., & Associates. (2000). *Learning as transformation: Critical perspectives on a theory in progress*. San Francisco, CA: Jossey-Bass.
- Michael, J. (2003). Using the Myers-Briggs Type Indicator as a tool for leadership development? Apply with caution. *Journal of Leadership & Organizational Studies*, 10(1), 68–81. <https://doi.org/10.1177/107179190301000106>
- Miles, M. B., & Huberman, M. A. (1994). *Qualitative data analysis: An expanded sourcebook*. Thousand Oaks, CA: Sage Publications Ltd.
- Miller, D. (1993). The architecture of simplicity. *Academy of Management Review*, 18(1), 116–138. <https://doi.org/10.5465/AMR.1993.3997509>

- Mintzberg, H., Lampel, J., Quinn, J. B., & Ghoshal, S. (2002). *The strategy process: Concepts, contexts, cases* (4th ed.). Upper Saddle River, NJ: Financial Times Management.
- Mischel, W. (1968). *Personality and assessment*. New York, NY: Wiley.
- Moore, T. (1987). Personality tests are back. *Fortune*, 115(7), 74.
- Morgan, A., Cannan, K., & Cullinane, J. (2005). 360 feedback: A critical enquiry. *Personnel Review*, 34(6), 663–680. <https://doi.org/10.1108/00483480510623457>
- Muir, D. (2014). Mentoring and leader identity development: A case study. *Human Resource Development Quarterly*, 25(3), 349–379. <https://doi.org/10.1002/hrdq.21194>
- Myers, I. B. (1962). *Manual: The Myers-Briggs Type Indicator*. Palo Alto, CA: Consulting Psychologists Press.
- Myers, I. B., Kirby, L. K., & Myers, K. D. (2000). *Introduction to type: A guide to understanding your results on the Myers-Briggs Type Indicator*. Palo Alto, CA: Consulting Psychologists Press.
- Naquin, S. S., & Holton, E. F. (2002). The effects of personality, affectivity, and work commitment on motivation to improve work through learning. *Human Resource Development Quarterly*, 13(4), 357–376. <https://doi.org/10.1002/hrdq.1038>
- Nelson, E., & Hogan, R. (2009). Coaching on the dark side. *International Coaching Psychology Review*, 4(1), 7–19.
- Nesbit, P. L. (2012). The role of self-reflection, emotional management of feedback, and self-regulation processes in self-directed leadership development. *Human Resource Development Review*, 11(2), 203–226. <https://doi.org/10.1177/1534484312439196>
- Newman, M. (2012). Calling Transformative Learning into question: Some mutinous thoughts. *Adult Education Quarterly*, 62(1), 36–55. <https://doi.org/10.1177/0741713610392768>
- Newton, B. J., Rothlingova, Z., Gutteridge, R., LeMarchand, K., & Raphael, J. H. (2011). No room for reflexivity? Critical reflections following a systematic review of qualitative research. *Journal of Health Psychology*, 17(6), 866–885. <https://doi.org/10.1177/1359105311427615>
- Nikolaou, I., & Oostrom, J. K. (Eds.). (2015). *Employee recruitment, selection, and assessment: Contemporary issues for theory and practice*. Hove, UK: Psychology Press.
- O'Neil, J., & Marsick, V. (2007). *Understanding Action Learning: Theory into practice*. New York, NY: American Management Association.
- Österreichisches Normungsinstitut. (2005). *ÖNORM D 4000 - Anforderungen an Prozesse und Methoden in der Personalauswahl und -entwicklung*. Wien: Fachnormenausschuss 249 des Österreichischen Normungsinstitutes.
- Park, C. L. (2010). Making sense of the meaning literature: an integrative review of meaning making and its effects on adjustment to stressful life events. *Psychological Bulletin*, 136(2), 257.

-
- Passmore, J. (2008). *Psychometrics in coaching: Using psychological and psychometric tools for development*. London, UK: Kogan Page Publishers.
- Passmore, J., Holloway, M., & Rawle-Cope, M. (2010). Using MBTI type to explore differences and the implications for practice for therapists and coaches: Are executive coaches really like counsellors? *Counselling Psychology Quarterly*, 23(1), 1–16. <https://doi.org/10.1080/09515071003679354>
- Paul, A. M. (2005). *The cult of personality testing: How personality tests are leading us to miseducate our children, mismanage our companies, and misunderstand ourselves*. New York, NY: Free Press.
- Peltier, J. W., Hay, A., & Drago, W. (2005). The reflective learning continuum: Reflecting on reflection. *Journal of Marketing Education*, 27(3), 250–263. <https://doi.org/10.1177/0273475305279657>
- Piaget, J. (1950). *The psychology of intelligence* (M. Piercy and D. Berlyne, trans.). New York, NY: Routledge.
- Pittenger, D. J. (2005). Cautionary comments regarding the Myers-Briggs Type Indicator. *Consulting Psychology Journal: Practice and Research*, 57(3), 210. <https://doi.org/10.1037/1065-9293.57.3.210>
- Poell, R. F. (2015). HRD Education in the Netherlands. *Advances in Developing Human Resources*, 17(2), 180–195. <https://doi.org/10.1177/1523422315572620>
- Poell, R. F., Lundgren, H., Bang, A., Justice, S. B., Marsick, V. J., Sung, S., & Yorks, L. (2018). How do employees' individual learning paths differ across occupations? A review of 10 years of empirical research. *Journal of Workplace Learning*, 30(5), 315–334. <https://doi.org/10.1108/JWL-01-2018-0019>
- Poell, R. F., & Van Der Krogt, F. (2017a). Why is organizing human resource development so problematic? Perspectives from the learning-network theory (Part I). *The Learning Organization*, 24(3), 180–193. <https://doi.org/10.1108/TLO-12-2016-0093>
- Poell, R. F., & Van Der Krogt, F. (2017b). Why is organizing human resource development so problematic? Perspectives from the learning-network theory (Part II). *The Learning Organization*, 24(4), 215–225. <https://doi.org/10.1108/TLO-12-2016-0094>
- Poell, R. F., & Van der Krogt, F. J. (2010). Individual learning paths of employees in the context of social networks. In S. Billet (Ed.), *Learning through practice* (pp. 197–221). London: Springer.
- Poell, R. F., & Van der Krogt, F. J. (2014). An empirical typology of hospital nurses' individual learning paths. *Nurse Education Today*, 34(3), 428–433. <https://doi.org/10.1016/j.nedt.2013.04.005>
- Poell, R. F., & Van der Krogt, F. J. (2015). The Learning-Network Theory: Actors organize dynamic HRD networks. In R. F. Poell, T. S. Rocco, & G. L. Roth (Eds.), *The Routledge Companion to Human Resource Development*. London: Routledge.

- Powell, J. H. (1989). The reflective practitioner in nursing. *Journal of Advanced Nursing*, 14(10), 824–832. <https://doi.org/10.1111/j.1365-2648.1989.tb01467.x>
- Pryor, R. G., & Bright, J. E. (2014). The Chaos Theory of Careers (CTC): Ten years on and only just begun. *Australian Journal of Career Development*, 23(1), 4–12. <https://doi.org/10.1177/1038416213518506>
- Raelin, J. A. (2001). Public reflection as the basis of learning. *Management Learning*, 32(1), 11–30. <https://doi.org/10.1177/1350507601321002>
- Raelin, J. A. (2002). “I don’t have time to think!” versus the art of reflective practice. *Reflections*, 4(1), 66–79. <https://doi.org/10.1162/152417302320467571>
- Reay, T., Berta, W., & Kohn, M. K. (2009). What’s the evidence on evidence-based management? *The Academy of Management Perspectives*, 23(4), 5–18. <https://doi.org/10.5465/amp.23.4.5>
- Rentfrow, P. J., Jokela, M., & Lamb, M. E. (2015). Regional Personality Differences in Great Britain. *PloS One*, 10(3), e0122245. <https://doi.org/10.1371/journal.pone.0122245>
- Richardson, G., & Maltby, H. (1995). Reflection-on-practice: Enhancing student learning. *Journal of Advanced Nursing*, 22(2), 235–242. <https://doi.org/10.1046/j.1365-2648.1995.22020235.x>
- Rihoux, B., & Ragin, C. C. (2009). *Configurational comparative methods: Qualitative comparative analysis (QCA) and related techniques* (Vol. 51). Thousand Oaks, CA: Sage Publications.
- Roberts, B. W., & Mroczek, D. (2008). Personality trait change in adulthood. *Current Directions in Psychological Science*, 17(1), 31. <https://doi.org/10.1111/j.1467-8721.2008.00543.x>
- Rogers, R. R. (2001). Reflection in higher education: A concept analysis. *Innovative Higher Education*, 26(1), 37–57. <https://doi.org/10.1023/A:1010986404527>
- Rouleau, L. (2005). Micro-practices of strategic sensemaking and sensegiving: How middle managers interpret and sell change every day. *Journal of Management Studies*, 42(7), 1413–1441. <https://doi.org/10.1111/j.1467-6486.2005.00549.x>
- Ryan, A. M., Brutus, S., Greguras, G. J., & Hakel, M. D. (2000). Receptivity to assessment-based feedback for management development. *Journal of Management Development*, 19(4), 252–276. <https://doi.org/10.1108/02621710010322580>
- Ryan, A. M., & Ployhart, R. E. (2000). Applicants’ perceptions of selection procedures and decisions: A critical review and agenda for the future. *Journal of Management*, 26(3), 565–606. [https://doi.org/10.1016/S0149-2063\(00\)00041-6](https://doi.org/10.1016/S0149-2063(00)00041-6)
- Salgado, J. F., & Tauriz, G. (2014). The Five-Factor Model, forced-choice personality inventories and performance: A comprehensive meta-analysis of academic and occupational validity studies. *European Journal of Work and Organizational Psychology*, 23(1), 3–30. <https://doi.org/10.1080/1359432X.2012.716198>

-
- Scharlau, C. (2004). Myers-Briggs Typenindikator - ein ressourcenfokussierendes Instrument für Coaching und Karriereberatung. *Organisationsberatung, Supervision, Coaching*, 11(1), 13–25. <https://doi.org/10.1007/s11613-004-0003-4>
- Schippers, M. C., Den Hartog, D. N., & Koopman, P. L. (2007). Reflexivity in teams: A measure and correlates. *Applied Psychology*, 56(2), 189–211. <https://doi.org/10.1111/j.1464-0597.2006.00250.x>
- Schmitt, N., & Chan, D. (1999). The status of research on applicant reactions to selection tests and its implications for managers. *International Journal of Management Reviews*, 1(1), 45–62. <https://doi.org/10.1111/1468-2370.00004>
- Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*. New York, NY: Basic Books.
- Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., ... others. (2013). Personality, gender, and age in the language of social media: The open-vocabulary approach. *PloS One*, 8(9), e73791. <https://doi.org/10.1371/journal.pone.0073791>
- Scully-Russ, E. (2005). Agency versus structure: Path dependency and choice in low-wage labor markets. *Human Resource Development Review*, 4(3), 254–278. <https://doi.org/10.1177/1534484305278349>
- Sfard, A. (1998). On two metaphors for learning and the dangers of choosing just one. *Educational Researcher*, 27(2), 4–13. <https://doi.org/10.3102/0013189X027002004>
- Sherlock, J., & Nathan, M. (2004). Producing actionable knowledge: Applying Mezirow's theory to the managerial learning context. In *Academy of Management Proceedings* (Vol. 2004, pp. E1–E6). New Orleans, LA: Academy of Management.
- Short, D. C. (2006). Closing the gap between research and practice in HRD. *Human Resource Development Quarterly*, 17(3), 343–350. <https://doi.org/10.1002/hrdq.1178>
- Short, D. C., & Yorks, L. (2002). Analyzing training from an emotions perspective. *Advances in Developing Human Resources*, 4(1), 80–96. <https://doi.org/10.1177/1523422302004001006>
- Siegel, H. (1988). *Educating reason: Rationality, critical thinking, and education*. New York, NY: Routledge.
- Smith, M., & Smith, P. (2005). *Testing people at work: Competencies in psychometric testing*. Malden, UK: BPS Blackwell Publishing.
- Snyder, C. (2008). Grabbing hold of a moving target: Identifying and measuring the Transformative Learning process. *Journal of Transformative Education*, 6(3), 159–181. <https://doi.org/10.1177/1541344608327813>
- Srivastava, S., John, O. P., Gosling, S. D., & Potter, J. (2003). Development of personality in early and middle adulthood: Set like plaster or persistent change? *Journal of Personality and Social Psychology*, 84(5), 1041. <https://doi.org/10.1037/0022-3514.84.5.1041>

- Stewart, J., Mills, S., & Sambrook, S. (2015). HRD programs in the United Kingdom. *Advances in Developing Human Resources, 17*(2), 162–179. <https://doi.org/10.1177/1523422315572619>
- Stothart, C. (2011). Coaching with 360 and MBTI: A case study. *Assessment and Development Matters, 3*(3).
- Sullivan, W. M. (2005). *Work and integrity: The crisis and promise of professionalism in North America* (2nd ed.). San Francisco, CA: Jossey-Bass Inc Pub.
- Swanson, G. (1971). Frameworks for comparative research: Structural anthropology and the theory of action. In I. Vallier & D. E. Apter (Eds.), *Comparative methods in sociology: Essays on trends and applications* (pp. 141–202). Berkeley, CA: University of California Press.
- Swanson, R. A., & Holton, E. F. (2001). *Foundations of human resource development*. San Francisco, CA: Berrett-Koehler Publishers.
- Talbot, M. (1999, October 17). The Rorschach chronicles. *New York Times Magazine*, pp. 28–31.
- Taylor, E. W. (1997). Building upon the theoretical debate: A critical review of the empirical studies of Mezirow's Transformative Learning Theory. *Adult Education Quarterly, 48*(1). <https://doi.org/10.1177/074171369704800104>
- Taylor, E. W. (1998). *The theory and practice of Transformative Learning: A critical review* (Information Series No. 374). Columbus, OH: ERIC Clearinghouse on Adult, Career, and Vocational Education.
- Taylor, E. W. (2007). An update of Transformative Learning Theory: A critical review of the empirical research (1999–2005). *International Journal of Lifelong Education, 26*(2), 173–191. <https://doi.org/10.1080/02601370701219475>
- Taylor, E. W., & Cranton, P. (2012). *The handbook of transformative learning: Theory, research, and practice*. San Francisco, CA: Jossey-Bass.
- Tenhiälä, A., Giluk, T. L., Kepes, S., Simón, C., Oh, I.-S., & Kim, S. (2016). The Research-Practice gap in human resource management: A Cross-Cultural study. *Human Resource Management, 55*(2), 179–200. <https://doi.org/10.1002/hrm.21656>
- test.de. (2014, July 9). Was bin ich? Persönlichkeitstests im Internet [Report]. Retrieved February 3, 2016, from <http://www.test.de/Persoenlichkeitstests-im-Internet-Was-bin-ich-4727586-0/>
- Tett, R. P., & Burnett, D. D. (2003). A personality trait-based interactionist model of job performance. *Journal of Applied Psychology, 88*(3), 500. <https://doi.org/10.1037/0021-9010.88.3.500>
- Tett, R. P., & Christiansen, N. D. (2007). Personality tests at the crossroads: A response to Morgeson, Campion, Dipboye, Hollenbeck, Murphy, and Schmitt (2007). *Personnel Psychology, 60*(4), 967–993. <https://doi.org/10.1111/j.1744-6570.2007.00098.x>
- Thomas, K. J. (2014). Workplace technology and the creation of boundaries: The role of VHRD in a 24/7 work environment. *Advances in Developing Human Resources, 16*(3), 281–295. <https://doi.org/10.1177/1523422314532092>

- Tisdell, E. J., & Tolliver, D. E. (2001). The role of spirituality in culturally relevant and transformative adult education. *Adult Learning, 12*(3), 13. <https://doi.org/10.1177/104515950101200305>
- Tobacyk, J., Milford, G., Springer, T., & Tobacyk, Z. (1988). Paranormal beliefs and the Barnum effect. *Journal of Personality Assessment, 52*(4), 737-739. https://doi.org/10.1207/s15327752jpa5204_13
- Tomlinson, H. (2004). *Educational leadership: Personal growth for professional development*. London, UK: Sage.
- Torraco, R. J., & Lundgren, H. (2018). *What HRD is doing - What HRD should be doing: The case for transforming HRD*. Working paper, University of Nebraska-Lincoln.
- van de Ven, N., Bogaert, A., Serlie, A., Brandt, M. J., & Denissen, J. J. (2017). Personality perception based on LinkedIn profiles. *Journal of Managerial Psychology, 32*(6), 418-429. <https://doi.org/10.1108/JMP-07-2016-0220>
- Van Woerkom, M. (2003). *Critical reflection at work: Bridging individual and organizational learning*. Enschede, NL: Twente University Press.
- Van Woerkom, M. (2004). The concept of critical reflection and its implications for human resource development. *Advances in Developing Human Resources, 6*(2), 178-192. <https://doi.org/10.1177/1523422304263328>
- Van Woerkom, M. (2008). Critical reflection and related higher-level conceptualizations of learning: Realistic or idealistic? *Human Resource Development Review, 7*(1), 3-12. <https://doi.org/10.1177/1534484307311804>
- Van Woerkom, M. (2010). Critical reflection as a rationalistic ideal. *Adult Education Quarterly, 60*(4), 339. <https://doi.org/10.1177/0741713609358446>
- Vermeren, P. (2009). *De HR-ballon - 10 populaire praktijken doorprikt* (2nd ed.). Gent, Belgium: Academia Press.
- Vince, R. (2002). Organizing reflection. *Management Learning, 33*(1), 63-78. <https://doi.org/10.1177/1350507602331003>
- Von Glasersfeld, E. (1984). An introduction to radical constructivism. In P. Watzlawick (Ed.), *The invented reality* (pp. 17-40). London, UK: W. W. Norton & Company.
- Wald, H. S., Borkan, J. M., Taylor, J. S., Anthony, D., & Reis, S. P. (2012). Fostering and evaluating reflective capacity in medical education: Developing the REFLECT rubric for assessing reflective writing. *Academic Medicine, 87*(1), 41. <https://doi.org/10.1097/ACM.0b013e31823b55fa>
- Wallman, A., Lindblad, A. K., Hall, S., Lundmark, A., & Ring, L. (2008). A categorization scheme for assessing pharmacy students' levels of reflection during internships. *American Journal of Pharmaceutical Education, 72*(1), Article 05. <https://doi.org/10.5688/aj720105>
- Watkins, K. (1989). Business and Industry. In S. B. Merriam & P. M. Cunningham (Eds.), *Handbook of adult and continuing education* (pp. 422-435). San Francisco, CA: Jossey-Bass.
- Weber, L., & Dwoskin, E. (2014, September 29). Are workplace personality tests fair? *Wall Street Journal*. Retrieved from <http://www.wsj.com/articles/areworkplacepersonalitytests-fair1412044257>

- Weick, K. E. (1995). *Sensemaking in organizations* (Vol. 3). Thousand Oaks, CA: Sage.
- Weick, K. E., Sutcliffe, K. M., & Obstfeld, D. (2005). Organizing and the process of sensemaking. *Organization Science*, 16(4), 409–421. <https://doi.org/10.1287/orsc.1050.0133>
- Wiggins-Romesburg, C. A., & Githens, R. P. (2018). The psychology of diversity resistance and integration. *Human Resource Development Review*, 17(2), 179–198. <https://doi.org/10.1177/1534484318765843>
- Williams, W. R. (2012). *Emergenetics profile* (Technical Report) (pp. 1–22). Centennial, CO: Emergenetics LLC.
- Wolff, H.-G., & Kim, S. (2012). The relationship between networking behaviors and the Big Five personality dimensions. *Career Development International*, 17(1), 43–66. <https://doi.org/10.1108/13620431211201328>
- Yanow, D., & Tsoukas, H. (2009). What is reflection-in-action? A phenomenological account. *Journal of Management Studies*, 46(8), 1339–1364. <https://doi.org/10.1111/j.1467-6486.2009.00859.x>
- Yin, R. K. (2004). *Case study research: Design and methods* (3rd ed.). Thousand Oaks, CA: Sage Publications Ltd.
- Yorks, L. (2004). Toward a political economy model for comparative analysis of the role of strategic human resource development leadership. *Human Resource Development Review*, 3(3), 189–208. <https://doi.org/10.1177/1534484304266260>
- Yorks, L., & Kasl, E. (2002). Toward a theory and practice for whole-person learning: Reconceptualizing experience and the role of affect. *Adult Education Quarterly*, 52(3), 176–192. <https://doi.org/10.1177/07417136020523002>
- Yorks, L., & Marsick, V. J. (2000). Organizational learning and transformation. In J. Mezirow (Ed.), *Learning as transformation: Critical perspectives on a theory in progress* (pp. 253–281). San Francisco, CA: John Wiley & Sons.
- Youyou, W., Kosinski, M., & Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. *Proceedings of the National Academy of Sciences*, 112(4), 1036–1040. <https://doi.org/10.1073/pnas.1418680112>

Appendices

Appendices

Appendix 1: Personality test names and abbreviations

	Abbr.	Personality Test Names	Category
1	NEO PI-R	Neuroticism, Extraversion, Openness Personality Inventory (revised)	
2	OPQ	Occupational Personality Questionnaire	
3	Hogan	Hogan Personality Inventory (HPI); Hogan Development Survey (HDS); Motives Values Preferences Inventory (MVPI)	Trait
4	FIRO-B	Fundamental Interpersonal Relations Orientation - Behavior	
5	Big Five	Reflector Big Five	
6	Drives	Management Drives	
7	MBTI	Myers-Briggs Type Indicator	
8	TMP	Team Management Profile	
9	DISC	Dominance (D), Inducement (I), Submission (S) and Compliance (C)	Type
10	Insights	Insights Discovery Preference Evaluator	
11	My Motivation	My Motivation Insights, also referred to as Spiral Dynamics	
12	Belbin	Belbin Team Roles	Other

Appendix 2: Data collection overview of multiple-case study¹⁷

Document	Type	Stakeholder	Organization	Country	Date
Doc001a	Telephone	Manager for Training and Business Development	Distributor, Training & Consulting	UK	17-28.02.12
Doc001b ¹⁸	Emails, brochure				
Doc002	Telephone	Employee, editorial team	Publisher	D	22.03.12
Doc003	Telephone	HR Director (retired)	Public organization	D	24.03.12
Doc004	Telephone	Director Product Management	Distributor, Training & Consulting	D	30.03.12
Doc005	Telephone	Employee, psychometric testing team	Association (Psychology)	UK	03.04.2012
Doc006	Telephone	Employee, customer service	Association (Personnel and Development)	UK	03.04.2012
Doc007	Emails, brochure	Different employees, test development	Publisher	D	03-05.4.12
Doc008	Telephone	Product Manager	Publisher	UK	05.04.12
Doc009	Telephone	Continuing Education Consultant	Adult Education School	A	12.04.12
Doc010	Face to face (F2F)	Director Research & Development	Publisher, Distributor, Training & Consulting	UK	13.04.12
Doc011	Direct Observation	Learning & Development Fair	Publishers, Distributors	UK	26.04.12

¹⁷ Additional materials for review but not for publication

¹⁸ The “b” indicates that a different type of data was collected from the same stakeholder.

Document	Type	Stakeholder	Organization	Country	Date
Doc012a	F2F	Trainer	Training & Consulting	D	07.05.12
Doc012b	Telephone				25.07.12
Doc013	F2F	Professional development advisor	Higher Education, HR Department	UK	29.05.12
Doc014	F2F	Psychologist	Publisher	UK	12.06.12
Doc015	F2F, brochure	Employee, sales team	Publisher	UK	12.06.12
Doc016	Group discussion	Employees, school leadership education	Teacher Training College	CH	14.05.13
Doc017	Group discussion	Employees, various departments	Higher education institution	UK	06.03.14
Doc018	Skype	Former consultant, now academic	Publisher	NL	12.05.15
Doc019a	Skype	Psychologist	Consulting	NL	23-31.06.15
Doc019b	Emails				
Doc020	Telephone	Psychologist	Consulting	NL	11.08.15
Doc021	Skype	HR Manager	Consumer goods	D	09.09.16
Doc022	Telephone	Psychologist	Consulting (prior: banking)	NL	14.09.16
Doc023	Telephone	HR Director	Industrial	D	14.09.16
Doc024	Skype	Senior expert / consultant	Automotive	D	14.09.16
Doc025	Telephone	HR Business Partner	Industrial	NL	26.09.16
Doc026	Skype	HR Director	Consumer goods	D	27.09.16

Document	Type	Stakeholder	Organization	Country	Date
Doc027	Telephone	Psychologist	Consulting (prior: automotive)	D	01.10.16
Doc028	Telephone	Psychologist	Coaching	NL	04.10.16
Doc029a	Skype	Educator	In transition	NL	04.10.16
Doc029b	Email				06.10.16
Doc030	Skype	Consultant	Different industries (prior: airline and financial)	UK	26.10.16
Doc031	Telephone	Managing Partner	Coaching	NL	27.10.16
Doc032	Telephone	Learning and Development Manager	Research institution	UK	27.10.16
Doc033	Email	Head of HR	Industrial	UK	10.10.- 13.11.16
Doc034	Telephone	VP Organization Development	Technology services	UK	04.11.16
Doc035	Email	Employee, service team	Association (Psychology)	UK	17.11.16

Appendix 3: *Interview guide practitioner study*

Thank you for participating in this interview. I am interested in your professional experience with personality tests in HRD.

1. Tell me about your use of personality tests in your own professional practice. [-> Tests in general, specific test use, areas of application, test takers]

- Where and when do you use them for development in your organization?
- Do you use any personality tests specifically? Why these?
- Who are at the receiving end of these personality tests? Who are your test takers?

2. How do you acquire those personality tests? [-> Publishers and intermediaries; factors that influence decision-making]

- Do you use intermediaries, like coaches or trainers who are accredited in administering those tests? Or: Do you work directly with publishers?
- How do test publishers and intermediaries influence you in choosing and using a specific test?
- What role does the _____ professional organization play in your decision-making? (*Check interview form*)

3. When you use these personality tests in HRD, how would you describe the main purpose? [-> Purpose, implementation success, satisfaction/dissatisfaction]

- How do you reason the use of a psychometric test?
- How do you know that you have achieved this purpose?

- How satisfied are you with the use of personality testing in this context?

4. Personality tests are often criticized for supporting “pigeonholing” rather than critical reflection and development. How do you see them?

[-> Criticism; dealing with criticism; reasons of non-use; repurposing]

- What are your concerns when using personality testing in HRD?
- How do you negotiate these concerns/dilemmas?
- Have you ever had a “difficult” test taker who argued the use of these tests?
How did you respond?

5. What other “tools” do you use to encourage (self-)reflection in the workplace? [-> Other tools; purpose; concerns]

- When and where do you use these other tools?
- And with what purpose?
- What needs, concerns and dilemmas do you have using those tools?

6. Is there anything else that you would like to add?

.....

About the author

Henriette Lundgren is an international scholar-practitioner in the field of human resource development (HRD) with an interest in adult education, organizational psychology and talent management. Educated in the Netherlands, Italy and Germany, Henriette holds a degree in International Business from Maastricht University, a master's in Organizational Psychology from the Open University Hagen, and a Ph.D. in Human Resource Studies from Tilburg University.

After college, Henriette joined Unilever where she worked in various line management positions in the UK, Poland, the Netherlands, and Germany. While working in Hamburg, she decided to take up a part-time study in organizational psychology, focusing her master's thesis on the motivation for continuing education. Upon completion, Henriette joined Grange Partnership in the UK, a global training consultancy where she worked as partner and learning consultant with clients in Asia, Europe and Latin America.

Henriette's path as a scholar-practitioner continued when she initiated her doctoral research at the University of Hamburg, Department of Adult Education, and later transferred to Tilburg University, Department of Human Resource Studies. Henriette's dissertation project focused on the purpose and perception of personality testing in the context of HRD.

The results of her Ph.D. are described in this book and were published in international peer-reviewed journals. Beyond her dissertation, Henriette engages in applied research in adult education and human resource studies, mainly through a research collaborative with Teachers College, Columbia University. After moving to the USA, Henriette continued her HRD career as the Director of Corporate Learning at Byrne Dairy. Henriette now lives with her husband in Upstate New York where she works as Talent Manager at Corning Incorporated.

Personality tests like the Myers–Briggs type indicator (MBTI) or Big Five assessments enjoy great popularity in business, human resource development (HRD), and adult education settings. However, the appeal of these tests to some is harshly criticized by others. In *Pigeonholing or Learning Instrument*, Henriette Lundgren poses central questions about how personality tests are being used in developmental settings, and to what extent their application can be perceived as ethical.

The argument is organized around four studies. The first study explores instruments, stakeholders, and practice dynamics in three Western European countries. The second study sets the theoretical stage with an in-depth literature review on reflection as operationalized by Jack Mezirow. The third study examines HRD practitioners and how they select their instruments. Finally, the fourth study analyzes managers' reactions to the use of these instruments as test takers. Throughout the book, the author reviews what can be learned by examining policy, professional identities, organizational sensemaking, and the larger societal marketplace dynamics that enable businesses to exploit such tensions around personality testing.

With this book, the author, herself an experienced HRD practitioner and researcher, offers new grounds and strategies for the use of personality testing in developmental contexts.