

Behavior-Based Assessment in Psychology

Tuulia M. Ortner
Fons J. R. van de Vijver
(Editors)

Going Beyond Self-Report
in the Personality, Affective,
Motivation, and Social Domains

Behavior-Based Assessment in Psychology

About the Editors

Tuulia M. Ortner, PhD, has been head of the Department of Psychological Assessment at the University of Salzburg, Austria, since 2012. She started working on objective personality tests more than 15 years ago at the University of Vienna, Austria, and continued her research at the Free University of Berlin, Germany. She is a member of the Executive Committee of the European Association of Psychological Assessment. Her current research includes examining the validity of behavior-based tests and their convergence with other measures.

Fons J. R. van de Vijver, PhD, is professor of Cross-Cultural Psychology at Tilburg University, The Netherlands, and holds honorary positions at North-West University, South Africa, and the University of Queensland, Australia. He has published widely on assessment issues, notably in the area of cross-cultural psychology, and also serves on the Executive Committee of the European Association of Psychological Assessment.

Psychological Assessment – Science and Practice

Each volume in the series *Psychological Assessment – Science and Practice* presents the state-of-the-art of assessment in a particular domain of psychology, with regard to theory, research, and practical applications. Editors and contributors are leading authorities in their respective fields. Each volume discusses, in a reader-friendly manner, critical issues and developments in assessment, as well as well-known and novel assessment

tools. The series is an ideal educational resource for researchers, teachers, and students of assessment, as well as practitioners.

Psychological Assessment – Science and Practice is edited with the support of the European Association of Psychological Assessment (EAPA).

Editor-in-Chief: Anastasia Efklides, Greece

Editorial Board: Itziar Alonso-Arbiol, Spain; Tuulia M. Ortner, Austria; Willibald Ruch, Switzerland; Fons J. R. van de Vijver, The Netherlands

Psychological Assessment – Science and Practice, Vol. 1

Behavior-Based Assessment in Psychology

Going Beyond Self-Report in the Personality, Affective, Motivation, and Social Domains

Edited by

Tuulia M. Ortner and Fons J. R. van de Vijver



Library of Congress Cataloging in Publication information for the print version of this book is available via the Library of Congress Marc Database under the LC Control Number 2015936493

Library and Archives Canada Cataloguing in Publication

Behavior-based assessment in psychology : going beyond self-report in the personality, affective, motivation, and social domains / edited by Tuulia M. Ortner and Fons J.R. van de Vijver.

(Psychological assessment—science and practice; vol. 1)

Includes bibliographical references and index.

Issued in print and electronic formats.

ISBN 978-0-88937-437-9 (paperback).—ISBN 978-1-61676-437-1 (pdf).—ISBN 978-1-61334-437-8 (html)

1. Behavioral assessment. I. Ortner, Tuulia M., author, editor II. Vijver, Fons J. R. van de, author, editor III. Series: Psychological assessment—science and practice; vol. 1

BF176.5.B44 2015

155.2'8

C2015-902319-X

C2015-902320-3

2015 © by Hogrefe Publishing

<http://www.hogrefe.com>

PUBLISHING OFFICES

USA: Hogrefe Publishing Corporation, 38 Chauncy Street, Suite 1002, Boston, MA 02111
Phone (866) 823-4726, Fax (617) 354-6875; E-mail customerservice@hogrefe.com

EUROPE: Hogrefe Publishing GmbH, Merkelstr. 3, 37085 Göttingen, Germany Phone +49 551 99950-0, Fax +49 551 99950-111; E-mail publishing@hogrefe.com

SALES & DISTRIBUTION

USA: Hogrefe Publishing, Customer Services Department, 30 Amberwood Parkway, Ashland, OH 44805

Phone (800) 228-3749, Fax (419) 281-6883; E-mail customerservice@hogrefe.com

UK: Hogrefe Publishing, c/o Marston Book Services Ltd., 160 Eastern Ave., Milton Park, Abingdon, OX14 4SB, UK

Phone +44 1235 465577, Fax +44 1235 465556; E-mail direct.orders@marston.co.uk

EUROPE: Hogrefe Publishing, Merkelstr. 3, 37085 Göttingen, Germany

Phone +49 551 99950-0, Fax +49 551 99950-111; E-mail publishing@hogrefe.com

OTHER OFFICES

CANADA: Hogrefe Publishing, 660 Eglinton Ave. East, Suite 119-514, Toronto, Ontario, M4G 2K2

SWITZERLAND: Hogrefe Publishing, Länggass-Strasse 76, CH-3000 Bern 9

Copyright Information

The e-book, including all its individual chapters, is protected under international copyright law. The unauthorized use or distribution of copyrighted or proprietary content is illegal and could subject the

purchaser to substantial damages. The user agrees to recognize and uphold the copyright.

License Agreement

The purchaser is granted a single, nontransferable license for the personal use of the e-book and all related files.

Making copies or printouts and storing a backup copy of the e-book on another device is permitted for private, personal use only.

Other than as stated in this License Agreement, you may not copy, print, modify, remove, delete, augment, add to, publish, transmit, sell, resell, create derivative works from, or in any way exploit any of the e-book's content, in whole or in part, and you may not aid or permit others to do so. You shall not: (1) rent, assign, timeshare, distribute, or transfer all or part of the e-book or any rights granted by this License Agreement to any other person; (2) duplicate the e-book, except for reasonable backup copies; (3) remove any proprietary or copyright notices, digital watermarks, labels, or other marks from the e-book or its contents; (4) transfer or sublicense title to the e-book to any other party.

These conditions are also applicable to any audio or other files belonging to the e-book. Should the print edition of this book include electronic supplementary material then all this material (e.g., audio, video, pdf files) is also available in the e-book edition.

Cover design: MetaDesign AG

Format: EPUB

ISBN 978-0-88937-437-9 (print) • ISBN 978-1-61676-437-1 (PDF) • ISBN 978-1-61334-437-8 (EPUB) <http://doi.org/10.1027/00437-000>

eBook-Herstellung und Auslieferung:
Brockhaus Commission, Kornwestheim
www.brocom.de

Table of Contents

Part I: Introduction

Chapter 1 Assessment Beyond Self-Reports

Tuulia M. Ortner and Fons J. R. van de Vijver

Part II: Modes and Theoretical Foundations

Chapter 2 Implicit Association Tests, Then and Now

*Marco Perugini, Giulio Costantini, Juliette Richetin, and
Cristina Zogmaister*

Chapter 3 A Model of Moderated Convergence Between Direct, Indirect, and Behavioral Measures of Personality Traits

*Manfred Schmitt, Wilhelm Hofmann, Tobias Gschwendner,
Friederike Gerstenberg, and Axel Zinkernagel*

Chapter 4 Narrative Content Coding

Michael Bender

Chapter 5 Beyond Projection: Performance-Based Assessment

Robert F. Bornstein

Part III: Measures

Chapter 6 Measuring Implicit Motives

Athanasios Chasiotis

Chapter 7 Measures of Affect

Martina Kaufmann and Nicola Baumann

Chapter 8 Implicit Measures of Attitudes

Colin Tucker Smith and Kate A. Ratliff

Chapter 9 Objective Personality Tests

Tuulia M. Ortner and René T. Proyer

Part IV: Domains of Application

Chapter 10 Indirect Measures in the Domain of Health Psychology

Reinout W. Wiers, Katrijn Houben, Wilhelm Hofmann, and Alan W. Stacy

Chapter 11 Indirect Measures in Forensic Contexts

Alexander F. Schmidt, Rainer Banse, and Roland Imhoff

Chapter 12 Implicit Measures in Consumer Psychology

Malte Friese and Andrew Perkins

Chapter 13 Observation of Intra- and Interpersonal Processes

Axel Schölmerich and Julia Jäkel

Contributors

Subject Index

[1] Part I

Introduction

[2][3] Chapter 1

Assessment Beyond Self-Reports

Tuulia M. Ortner¹ and Fons J. R. van de Vijver²

¹Department of Psychology, University of Salzburg, Austria

²Department of Culture Studies, Tilburg University, The Netherlands

Self-reports have come under renewed scrutiny in the last few decades. Notably in social psychology, but spreading out to differential psychology, psychological assessment, and a number of applied fields of psychology, there is a tendency to refrain from using self-reports to collect data. This has led to a renewed interest in alternative modes of assessment. Objective measures or behavior-based measures are an example of such a method in which there is more interest than ever before, even though they have a long tradition, as shown here. This book provides an overview of the current state of the art in this field of assessment. In this introductory chapter we first give a short historical overview of the field, including a delineation of what we mean by *assessment beyond self-reports*. We then proceed by briefly describing the theme of each chapter. We conclude the chapter by drawing conclusions about the state of the field and its outlook.

A Short Look Back Into History

In the history of psychological assessment, behavior-based approaches for the measurement of personality characteristics and related constructs have played a major role from the very beginning. Early ancestors of personality psychology saw the relevance of behavioral indicators; examples are James McKeen Cattell, who in 1890 proposed behavioral tasks in his battery of

mental tests, and Francis Galton, who in 1884 stated that the measurement of aspects of character deserves carefully recorded acts. Later, leading scholars of human personality also included behavioral data into their research. For example, Raymond Bernard Cattell and his team proposed three sources of information in their integral assessment of personality including so-called T-data (referring to reactions to standardized experimental situations, besides L-data and Q-data, which involve everyday behaviors and self-reported questionnaire data, respectively) represented in measurement by so-called cursive miniature situations (Cattell, 1941,1944), later called objective tests. These tests aimed to stimulate the behavioral expression of personality while meeting common psychometric standards of psychological tests. Further earlier approaches of behavior-based assessment could be traced back to the early attempts of Herrmann Rorschach and his idea to interpret reactions to a set of ambiguous stimuli to refine clinical diagnoses by tapping into not explicitly verbalized aspects of personality (Rorschach, 1921).

[4] Nowadays, the available behavior-based approaches for the measurement of personality, motivational variables, or constructs addressing aspects related to social behavior represent an impressive variety of methods. This variety precludes a clear definition. Therefore, we refer to these as measurement approaches beyond self-reports. Such approaches *beyond self-reports* include the *basic* form of behavior observation and coding methods (e.g., index systems, category systems) that were found to be especially useful in the assessment and investigation of interactions (e.g., Hill, Maskowitz, Danis, & Wakschlag, 2008; Reyna, Brown, Pickler, Myers, & Younger, 2012), personality in children and adolescents (Kilgus, Riley-Tillman, Chafouleas, Christ, & Welsh, 2014; Martin-Storey, Serbin, Stack, & Schwartzman, 2009), and in the context of work and aptitude testing (Hennessy, Maybe, & Warr, 1998; Schollaert & Lievens, 2012). Measures beyond self-reports may also include analyses of the consequences of persons' behavior, such as the investigation of the abrasion of the floor in a museum in order to analyze visitors' preferences (unobtrusive measures;

Webb, Campbell, Schwartz, & Sechrest, 1966), or analyses of personal marks on the Internet, such as information given or activities conducted in social networks (Back et al., 2010). Also narratives, spoken or written statements or stories, represent written or recorded behavior and may serve as a source for personality assessment, using structured methods of content coding (Fiese & Spagnola, 2005; Kuefner, Back, Nestler, & Egloff, 2010). Furthermore, the use of psychophysiological measures as indicators of physiological arousal (e.g., Gannon, Beech, & Ward, 2008; Madsen, Parsons, & Grubin, 2004) or facial expressions as indicators of emotions (polygraph; Tracy, Robins, & Schriber, 2009; Vick, Waller, Parr, Pasqualini, & Bard, 2007) would fall into this category.

Measurement approaches beyond self-reports also include classic projective techniques (Lilienfeld, Wood, & Garb, 2000) that assess persons' responses to ambiguous stimuli. Their validity has been widely discussed in the last few decades (Bornstein, 1999; Viglione, 1999; Weiner, 1997) and newer developments, such as a semiprojective test, have been proposed with the aim of overcoming criticism leveled at projective tests, such as a shortage of objectivity in scoring and lack of interpretation of the scores based on normative samples (Sokolowski, Schmalt, Langens, & Puca, 2000). New computerized technologies further enabled a large number of testing procedures. The fledgling field is quickly growing, as demonstrated by a large number of new computerized objective personality tests building on Cattell's notion of the miniature situation (see Ortner & Schmitt, 2014) as well as by widely applied so-called indirect tests, mainly represented by reaction time measures (e.g., De Houwer, 2003; Greenwald, McGhee, & Schwartz, 1998; Payne, Cheng, Govorun, & Steward, 2005), but also including further indicators for indirect attitudes, such as evaluative decisions (Payne et al., 2005).

When addressing assessment instruments, procedures in the noncognitive domain (i.e., personality, affect, attitudes, and motivation), self-report questionnaires represent the dominant approach. All the behavior-based approaches mentioned are – compared with questionnaires assessing self-

reports – much less frequently employed in most domains of psychological research (see Alonso-Arbiol & Van de Vijver, 2010; Ortner & Vormittag, 2011) and practice (Evers et al., 2012). Why are these approaches less visible, less used, and less within the focus of research compared with self-reports? As far as behavior observations, narratives, and most projective techniques are concerned, one of the main reasons may be the effort involved in collecting and processing behavioral observations to assess persons' characteristics. Most behavior-based approaches of assessment produce much more data than questionnaires – data that need to be sorted, integrated, or summarized. Thus, test economy and procedural efforts may often be the reason to refrain from using these methods. However, this disadvantage does not apply to newer computerized indirect or objective testing procedures. The new technology may have led to their increased visibility and impact in current research.

[5] This volume is based on the premise that behavior-based assessment represents an essential element in the assessment process and should be included whenever possible. We propose the following reasons: First and foremost, objective measures suffer less or not at all from various well-documented problems of self-reports, such as response styles (e.g., Linden, Paulhus, & Dobson, 1986; Podsakoff & Organ, 1986) and the limitations of introspection (Howe, 1991; Nisbett & Wilson, 1977). For example, not all processes of interest in assessment can be accessed, remembered, and reported. Persons differ in their ability to identify real-life situations that are relevant to estimate certain constructs via self-reports and to integrate this information into a self-related judgment. Second, the nature and detail of assessed real behavior greatly exceed those of reported or estimated behavior. As the saying goes, actions speak louder than words. Third, researchers and practitioners do not have to pit one method against another by following the recommendation to use multiple methods in assessing a given construct in order to receive a more complete picture and to compensate the weaknesses that are inherent to specific measurement approaches (Fernandez-Ballesteros et al., 2001).

Chapters of the Book

In this volume we aim to address behavior-based assessment from researchers' and developers' perspective, up to its implementation in practice. The volume is divided into four parts. After this short introduction ([Part I, Chapter 1](#)), the second part ([Part II](#)) of the volume addresses particular modes of behavior-based assessment embedded in theoretical foundations. The first chapter of this part, [Chapter 2](#), by Marco Perugini, Giulio Costantini, Juliette Richetin, and Cristina Zogmaister, presents an introduction to the Implicit Association Test (IAT) as the most prominent representative of indirect measures today. The authors provide a definition of indirect measures, discuss cognitive processes underlying the IAT effect, and address its psychometric aspects by discussing the scoring of the IAT and its reliability and validity. [Chapter 3](#) by Manfred Schmitt, Wilhelm Hofmann, Tobias Gschwendner, Friederike Gerstenberg, and Axel Zinkernagel describes a new and innovative theoretical model. In line with the theory of planned behavior (Ajzen, 1987) and with the reflective impulsive model (RIM; Strack & Deutsch, 2004) they differentiate between manifest behavior, behavioral plans and intentions, and behavioral schemata or scripts. In their chapter, they postulate that the degree of convergence between direct, indirect, and behavioral measures is variable, not constant, and they propose a number of variables that moderate the convergence between the components of the model. Michael Bender gives an overview of thematic vs. structural analyses of texts and discusses these procedures and their usability in [Chapter 4](#). He further addresses a number of practical areas of application, such as analyses of autobiographic narratives, eyewitness reports, and the assessment of depression. Robert Bornstein takes the reader to a journey into the theory and practice of the Rorschach Inkblot Method (RIM) as a representative of the huge family of projective techniques in [Chapter 5](#). He addresses processes underlying Rorschach responses and discusses psychometric properties of this approach. His chapter closes with explicit guidelines for clinicians and clinical researchers for the use of RIM data.

Part III of this volume is dedicated to specific measures. The chapters in this part provide an introduction and overview on background information, psychometric properties, and recent developments of particular groups of measures. First, Athanasios Chasiotis presents different approaches to the measurement of implicit motives in **Chapter 6**. After an introduction into implicit and explicit motives, he presents the Picture Story Exercise (PSE) and the Operant^[6] Motive Test (OMT) as content-coding methods for the assessment of implicit motives. He discusses theoretical foundations, practical aspects of presentation and scoring, as well as their psychometric properties. Behavior-based methods for the assessment of affect are summarized and presented by Martina Kaufmann and Nicola Baumann in **Chapter 7**. They systematically address particular measures assigned to three groups of methods: indirect, reaction time-based approaches; projective techniques; and behavioral observations to assess affect. They discuss the possibilities and limitations of the approaches. Colin Tucker Smith and Kate Ratliff present an overview of indirect measures for the assessment of attitudes and their psychometric properties in **Chapter 8**, such as different variations of the IAT, the Evaluative Priming Task, the Go/No-Go Association Task, the Extrinsic Affective Simon Task (EAST), the Sorting Paired Features Task, and the Affective Misattribution Procedure (AMP). In the next chapter (**Chapter 9**), Tuulia Ortner and Rene Proyer give an overview of tests that derive personality-related characteristics from observable behavior on performance tasks or other highly standardized miniature situations that lack face validity, so-called objective personality tests (OPTs). As an attempt to group this heterogeneous group of tests, they introduce three categories of different OPTs: (a) OPTs masked as achievement tasks, (b) OPTs that aim to represent real-life simulations, and (c) questionnaire-type OPTs that ask for evaluations or decisions, but lack face validity since different constructs than suggested are assessed. Psychometric properties are addressed by giving a number of examples of contemporary OPTs. The chapter closes with an analysis of the current state in research and practice.

[Part IV](#) provides insight into approaches, methods, and empirical findings with reference to specific areas of practical application. Reinout Wiers, Katrijn Houben, Wilhelm Hofmann, and Alan W. Stacy discuss indirect measures in the domain of health psychology in [Chapter 10](#). They argue that initial, impulsive reactions, assessed by indirect measures, may be the most important predictor of health behaviors in some people in some situations. They introduce an impressive variety of measures and discuss their correlations in the health domain. In addition, they discuss the assessment of reflective processes. In [Chapter 11](#), Alexander Schmidt, Rainer Banse, and Roland Imhoff give an overview of indirect measures in a forensic context with special attention to the assessment of deviant sexual interest. They present a large number of so-called *task-relevant* and *task-irrelevant* measures and carefully discuss empirical findings and psychometric properties of these measures. They complete their chapter with an outlook on the future with reference to methodological aims, theoretical demands, and aims with regard to clinical implementation of indirect assessment. Behavior-based approaches within consumer psychology are discussed by Malte Friese and Andrew Perkins in [Chapter 12](#). They first present precursors of implicit measures and later provide an extensive review of empirical studies employing implicit measures in the consumer context. Finally, they provide an outlook and discuss some challenges for future research. In [Chapter 13](#), Axel Schölmerich and Julia Jäkel present advantages and challenges of observational methods (OM) for the assessment of intra- and interpersonal processes. After an introduction into behavior observation systems, they discuss several specific behavior observation instruments and their psychometric properties. In their conclusion, they evaluate procedures for behavior observation and formulate demands for their future development.

What Can We Learn From the Chapters?

In our view, the current chapters provide the basis for the following conclusions about the current state of behavior-based assessment:

- [7] 1. *The concrete relevance of behavior-based approaches depends on the context.* Various chapters clearly suggest that behavior-based approaches are most suitable in specific settings. For example, Robert Bornstein concludes in his chapter on current approaches for the use of projective techniques (Chapter 5) that the exclusive reliance on questionnaires assessing self-report is particularly critical in *clinical settings*, where self-reports of traits and symptoms reflect people's tendencies to view and (or) present themselves. A critical factor that should raise the interest in and relevance of behavior-based measures is lack of insight into characteristics of the construct being assessed (e.g., personality pathology). Another critical factor can be found in forensic psychology, where questionnaires and interviews are transparent and can easily be faked by respondents who are aware of the personal consequences of the assessment outcome; here, indirect approaches seem promising (Chapter 11). In health psychology, impulsive reactions captured by indirect approaches may be the most important predictor of health behaviors in some situations in some persons (Chapter 10). In other domains, such as consumer psychology, the attitudes of interest are not necessarily less accessible through self-reports, but researchers assume that indirect measures can nevertheless contribute in a meaningful way to the investigation of concepts and processes beyond self-reports (Chapter 12). We conclude that a number of different reasons may contribute to the inclusion of behavior-based approaches in different fields of application.
2. *Findings on the psychometric properties of one behavior-based measure cannot be generalized to another.* This means especially that reliability and validity need to be empirically examined and proven for each test or diagnostic procedure separately. This even means for most approaches in this volume that the same procedure, such as an IAT (or a behavior observation scheme, a narrative coding system, an OPT) that aims to assess one construct may be valid, whereas another IAT (or another

behavior observation scheme, another narrative coding system, another OPT) that aims to assess another construct may not be (see, e.g., [Chapter 2](#)). We know from research on questionnaires that the usefulness of instruments critically depends on the stimuli used (or technical procedure implemented, data interpreted) and their suitability to evoke and therefore measure a certain construct.

3. *Not all behavior-based approaches are convenient to validly measure all constructs or all possible aspects of a construct.* Each of the presented approaches is more or less suitable to assess certain constructs or particular aspects of a construct – and not to assess all possible constructs or attitudes. For example, indirect approaches in general have proved to be more able to assess implicit aspects of attitudes ([Chapters 2 and 8](#)). As referred to by Ortner and Proyer ([Chapter 9](#)), interpersonal behavior and personality variables (e.g., extraversion) may not be validly assessed through computerized miniature situations as represented by OPTs, but they may be very validly assessed through behavior observation. It may be more difficult on the other hand to assess introspective processes or evaluations through behavior observations. This means that the valid assessment of a certain construct or attitude of interest is often inseparably bound to one or several methods of measurement.
4. *More research is needed.* The status of knowledge significantly differs between the approaches. The currently available body of scientific knowledge available is strong for some behavior-based approaches, and weaker for others. The Web of Knowledge indicated 25,288 journal entries including the keyword *behavior observation* in July 2014, 5,972 entries for *projective technique or projective test*, 3,551 journal entries for the keyword *implicit association test*, 219 results for *narrative content coding*, and 32 publications listed for the combined keywords of *objective personality tests* and *Cattell*. However, most research in the social and behavioral sciences is still based on self-reports; the corpus of knowledge regarding behavior-based approaches is widely behind the current research available on self-report questionnaires.

- [8] 5. *Construct validity of behavior-based measures remains a challenge for future research.* As referred to by Schmitt et al. ([Chapter 3](#)), the construct validity of OPTs needs to be investigated by going beyond the traditional strategy of convergent and discriminant validation as employed in the multitrait-multimethod framework proposed by Campbell and Fiske (1959). The low convergence of certain behavior-based measures with questionnaires addressing self-reports with simultaneously demonstrated criterion validity deserves a new theoretical framework to explain and interpret the convergence or lack thereof. Cronbach and Meehl (1955) argued that a test's construct validity is given when empirical data confirm claims that were made based on a theory describing the given construct. The model proposed by Schmitt et al. in this volume ([Chapter 3](#)) postulates in line with dual-process theories (Strack & Deutsch, 2004) that explicit dispositions can be measured directly with self-report scales and that implicit dispositions can only be measured indirectly with procedures like the IAT. They further propose that explicit dispositions affect behavior via plans and intentions, and that implicit dispositions affect behavior via the automatic activation of behavioral scripts and schemata. This model goes beyond classic dual-process theories by assuming that these effects are moderated by personality and situation factors. In order to meet particular new challenges, inclusion of moderators of convergence in designs of validity studies may increase the convergence and indicate their validity more thoroughly compared with bare correlation coefficients. Nevertheless, the field is also open for further theoretical frameworks and developments.
6. *Reliability of (some) behavior-based measures needs further attention.* As referred to in several chapters, reliabilities for some behavior-based measures are low. For example, reliabilities differ widely across objective tests (see [Chapter 9](#)) and implicit measures (see [Chapter 11](#)). Besides, low reliabilities impact on correlations among measures and lead to difficulties in replicating findings (LeBel & Paunonen, 2011); low retest correlations may also, but not necessarily, indicate a higher

amount of state variance assessed compared with trait variance (e.g., Koch, Ortner, Eid, & Schmitt, 2014; Schmukle & Egloff, 2004). Nevertheless, early studies revealed that behavior is more inconsistent than self-reported attitudes are (Hartshorne & May, 1928; Mischel, 1968; Ross & Nisbett, 1991). Therefore, even substantial efforts in test design and scoring may not raise the reliability of behavior-based measures to levels that are known from self-report questionnaires. We may therefore need to adjust our views on the reliability of behavior-based measures.

7. *There is ample room for further developments in behavior-based assessment.* Newer indirect methods, such as the IAT, have triggered an amazing interest in psychological research, as described in several chapters of this volume. The IAT in particular is a procedure that has been thoroughly investigated with reference to its functioning and, as Perugini and colleagues report, how to best develop and use it. Nevertheless, the IAT is not a task that could be implemented in order to assess individuals' characteristics or make reliable comparisons between individuals. Due to its psychometric properties, it is still a measure for the assessment of attitudes of groups instead of individuals. Perugini and colleagues point out that there is substantial room for improvements within the paradigm itself and make suggestions for future improvements of IATs. Future developments are also expected both in the currently underinvestigated field of OPTs and in all fields of application, where the use of behavior-based approaches is still underrepresented.

[9] **Coda**

We hope that the publication of this book will enhance the understanding of behavior-based assessment and stimulate research on the topic. We would also like to encourage practitioners to use multimethod assessment by including various sources of information in the assessment process. We

believe that we can only understand the complexity of human behavior by combining various theoretical and assessment perspectives. Behavior-based measures and their underlying models have an important role to play in this endeavor.

Acknowledgments

The editors gratefully acknowledge Karin C. Berkhout's skillful assistance in the editorial process.

References

- Ajzen, I. (1987). Attitudes, traits, and actions: Dispositional prediction of behavior in personality and social psychology. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 20, pp. 1–63). New York, NY: Academic Press.
- Alonso-Arbiol, I., & Van de Vijver, F. J. R. (2010). A historical analysis of the European Journal of Psychological Assessment: A comparison of the earliest (1992-1996) and the latest years (2005-2009). *European Journal of Psychological Assessment, 26*, 238–247. <http://doi.org/10.1027/1015-5759/a000032>
- Back, M. D., Stopfer, J. M., Vazire, S., Gaddis, S., Schmukle, S. C., Egloff, B., & Gosling, S. D. (2010). Facebook profiles reflect actual personality, not self-idealization. *Psychological Science, 21*, 372–374. <http://doi.org/10.1177/0956797609360756>
- Bornstein, R. F. (1999). Criterion validity of objective and projective dependency tests: A meta-analytic assessment of behavioral prediction. *Psychological Assessment, 11*, 48–57. <http://doi.org/10.1037/1040-3590.11.1.48>
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81–105. <http://doi.org/10.1037/h0046016>
- Cattell, J. M. (1890). Mental tests and measurements. *Mind, 15*, 373–381. <http://doi.org/10.1093/mind/os-XV.59.373>
- Cattell, R. B. (1941). An objective test of character-temperament I. *Journal of General Psychology, 25*, 59–73. <http://doi.org/10.1080/00221309.1941.10544704>
- Cattell, R. B. (1944). An objective test of character-temperament II. *Journal of Social Psychology, 19*, 99–114. <http://doi.org/10.1080/00224545.1944.9918805>
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281–302. <http://doi.org/10.1037/h0040957>
- De Houwer, J. (2003). The Extrinsic Affective Simon Task. *Experimental Psychology and Health, 50*, 77–85. <http://doi.org/10.1026//1618-3169.50.2.77>
- Evers, A., Muniz, J., Bartram, D., Boben, D., Egeland, J., Fernandez-Hermida, J. R., ... Urbanek, T. (2012). Testing practices in the 21st century developments and European psychologists'

- opinions. *European Psychologist*, 17, 300–319. <http://doi.org/10.1027/1016-9040/a000102>
- Fernandez-Ballesteros, R., De Bruyn, E. E. J., Godoy, A., Hornke, L. F., Ter Laak, J., Vizcarro, C., ... Zaccagnini, J. L. (2001). Guidelines for the assessment process (GAP): A proposal for discussion. *European Journal of Psychological Assessment*, 17, 187–200. <http://doi.org/10.1027//1015-5759.17.3.187>
- Fiese, B. H., & Spagnola, M. (2005). Narratives in and about families: An examination of coding schemes and a guide for family researchers. *Journal of Family Psychology*, 19, 51–61. <http://doi.org/10.1037/0893-3200.19.1.51>
- [10] Galton, F. (1884). Measurement of character. *Forthnightly Review*, 36, 179–185.
- Gannon, T. A., Beech, A. R., & Ward, T. (2008). Does the polygraph lead to better risk prediction for sexual offenders? *Aggression and Violent Behavior*, 13, 29–44. <http://doi.org/10.1016/j.avb.2007.08.001>
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. K. L. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, 74, 1464–1480. <http://doi.org/10.1037/0022-3514.74.6.1464>
- Hartshorne, H., & May, M. A. (1928). *Studies in the nature of character: Studies in deceit (1)*. New York, NY: Macmillan.
- Hennessy, J., Maybe, B., & Warr, P. (1998). Assessment Centre Observation Procedures: An experimental comparison of traditional, checklist and coding methods. *International Journal of Selection and Assessment*, 6, 222–231. <http://doi.org/10.1111/1468-2389.00093>
- Hill, C., Maskowitz, K., Danis, B., & Wakschlag, L. (2008). Validation of a clinically sensitive, observational coding system for parenting behaviors: The Parenting Clinical Observation Schedule. *Parenting-Science and Practice*, 8, 153–185. <http://doi.org/10.1080/15295190802045469>
- Howe, R. B. (1991). Introspection: A reassessment. *New Ideas in Psychology*, 9, 25–44. [http://doi.org/10.1016/0732-118X\(91\)90038-N](http://doi.org/10.1016/0732-118X(91)90038-N)
- Kilgus, S. P., Riley-Tillman, T. C., Chafouleas, S. M., Christ, T. J., & Welsh, M. E. (2014). Direct behavior rating as a school-based behavior universal screener: Replication across sites. *Journal of School Psychology*, 52, 63–82. <http://doi.org/10.1016/j.jsp.2013.11.002>
- Koch, T., Ortner, T. M., Eid, M., & Schmitt, M. (2014). Evaluating the construct validity of Objective Personality Tests using a Multitrait-Multimethod-Multioccasion (MTMM-MO) approach. *European Journal of Psychological Assessment*, 30, 208–230. <http://doi.org/10.1027/1015-5759/a000212>
- Kuefner, A. C. P., Back, M. D., Nestler, S., & Egloff, B. (2010). Tell me a story and I will tell you who you are! Lens model analyses of personality and creative writing. *Journal of Research in Personality*, 44, 427–435. <http://doi.org/10.1016/j.jrp.2010.05.003>
- LeBel, E. P., & Paunonen, S. V. (2011). Sexy but often unreliable: The impact of unreliability on the replicability of experimental findings with implicit measures. *Personality and Social Psychology Bulletin*, 37, 570–583. <http://doi.org/10.1177/0146167211400619>
- Lilienfeld, S. O., Wood, J. M., & Garb, H. N. (2000). The scientific status of projective techniques. *Psychological Science in the Public Interest*, 1(2), 27–66.
- Linden, W., Paulhus, D. L., & Dobson, K. S. (1986). Effects of response styles on the report of psychological and somatic distress. *Journal of Consulting and Clinical Psychology*, 54, 309–313. <http://doi.org/10.1037/0022-006X.54.3.309>

- Madsen, L., Parsons, S., & Grubin, D. (2004). A preliminary study of the contribution of periodic polygraph testing to the treatment and supervision of sex offenders. *Journal of Forensic Psychiatry & Psychology, 15*, 682–695. <http://doi.org/10.1080/1478994042000270256>
- Martin-Storey, A., Serbin, L. A., Stack, D. M., & Schwartzman, A. E. (2009). The behaviour style observation system for young children predicts teacher-reported externalizing behaviour in middle childhood. *Infant and Child Development, 18*, 337–350. <http://doi.org/10.1002/icd.601>
- Mischel, W. (1968). *Personality and assessment*. New York, NY: Wiley.
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*, 231–259. <http://doi.org/10.1037/0033-295X.84.3.231>
- Ortner, T. M., & Schmitt, M. (2014). Advances and continuing challenges in objective personality testing. *European Journal of Psychological Assessment, 30*, 163–168. <http://doi.org/10.1027/1015-5759/a000213>
- Ortner, T. M., & Vormittag, I. (2011). Articles published in EJPA 2009-2010: An analysis of the features of the articles and the characteristics of the authors. *European Journal of Psychological Assessment, 27*, 290–298. <http://doi.org/10.1027/1015-5759/a000082>
- Payne, B. K., Cheng, C. M., Govorun, O., & Steward, B. D. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology, 89*, 277–293. <http://doi.org/10.1037/0022-3514.89.3.277>
- Podsakoff, P. M., & Organ, D. W. (1986). Self-reports in organizational research: Problems and prospects. *Journal of Management, 12*, 531–544. <http://doi.org/10.1177/014920638601200408>
- [11] Reyna, B. A., Brown, L. F., Pickler, R. H., Myers, B. J., & Younger, J. B. (2012). Mother-infant synchrony during infant feeding. *Infant Behavior & Development, 35*, 669–677. <http://doi.org/10.1016/j.inf-beh.2012.06.003>
- Rorschach, H. (1921). *Psychodiagnostik. Der Rorschach-Test* [Psychodiagnostics. The Rorschach Test]. Bern, Switzerland: Huber.
- Ross, L., & Nisbett, R. E. (1991). *The person and the situation: Perspectives of social psychology*. New York, NY: McGraw-Hill.
- Schmukle, S. C., & Egloff, B. (2004). Does the implicit association test for assessing anxiety measure trait and state variance? *European Journal of Personality, 18*, 483–494. <http://doi.org/10.1002/per.525>
- Schollaert, E., & Lievens, F. (2012). Building situational stimuli in assessment center exercises: Do specific exercise instructions and role-player prompts increase the observability of behavior? *Human Performance, 25*, 255–271. <http://doi.org/10.1080/08959285.2012.683907>
- Sokolowski, K., Schmalt, H. D., Langens, T. A., & Puca, R. M. (2000). Assessing achievement, affiliation, and power motives all at once: The Multi-Motive Grid (MMG). *Journal of Personality Assessment, 74*, 126–145. <http://doi.org/10.1207/S15327752JPA740109>
- Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review, 8*, 220–247. http://doi.org/10.1207/s15327957pspr0803_1
- Tracy, J. L., Robins, R. W., & Schriber, R. A. (2009). Development of a FACS-verified set of basic and self-conscious emotion expressions. *Emotion, 9*, 554–559. <http://doi.org/10.1037/a0015766>
- Vick, S.-J., Waller, B. M., Parr, L. A., Pasqualini, M. C. S., & Bard, K. A. (2007). A cross-species comparison of facial morphology and movement in humans and chimpanzees using the Facial

- Action Coding System (FACS). *Journal of Nonverbal Behavior*, 31, 1–20. <http://doi.org/10.1007/s10919-006-0017-z>
- Viglione, D. J. (1999). A review of recent research addressing the utility of the Rorschach. *Psychological Assessment*, 11, 251–265. <http://doi.org/10.1037/1040-3590.11.3.251>
- Webb, E. J., Campbell, D. T., Schwartz, R. D., & Sechrest, L. (1966). *Unobtrusive measures: Nonreactive research in the social sciences*. Chicago, IL: Rand McNally.
- Weiner, I. B. (1997). Current status of the Rorschach Inkblot Method. *Journal of Personality Assessment*, 68, 5–19. http://doi.org/10.1207/s15327752jpa6801_2

[12][13] **Part II**

**Modes and Theoretical
Foundations**

[14][15] Chapter 2

Implicit Association Tests, Then and Now

Marco Perugini, Giulio Costantini, Juliette Richetin, and Cristina Zogmaister

Department of Psychology, University of Milan-Bicocca, Italy

One of the ways to understand the importance of a scientific contribution is by looking at how many times it is cited in the scientific literature. The original paper by Greenwald, McGhee, and Schwartz (1998) that presented the Implicit Association Test (IAT), published in the *Journal of Personality and Social Psychology* (JPSP), has so far been cited 1,900 times (as retrieved from Web of Science, March 26, 2012). Putting this figure into perspective, it is the most cited paper published in JPSP, the second most cited being a subsequent paper by Greenwald and colleagues on an improved scoring algorithm of the IAT (Greenwald, Nosek, & Banaji, 2003), and the fifth most cited paper in the whole field of psychology between 1998 and 2012. There is therefore little doubt that the IAT represents one of the most important developments in the field of psychology during the last 15 years. In this chapter we will first define direct and indirect measures, then present the IAT, discuss some cognitive processes behind its functioning, and briefly review some variants that have appeared in recent years. Adopting a psychometric perspective, the second part of this chapter will deal with issues such as the scoring of the IAT and its reliability and validity. The last part will focus on methodological issues relative to the development and the use of an IAT in a research context.

Throughout the chapter our review will provide an overview of what has been done (*then*), what is the current state of knowledge (now), and what are the potential interesting developments (*future*).

Direct and Indirect Measures

In this chapter we use the terms *direct* and *indirect* to refer to the measures, and explicit and implicit to refer to the constructs. We should, however, clarify that we have modified the definitions provided by De Houwer and Moors (2010). According to the authors:

...direct measures are characterized by two properties: (1) The measurement outcome is derived from a self-assessment by the participant. (2) The target of the self-assessment is the attribute that the measurement outcome is assumed to capture. If a measure does not have both of these properties, it can be called indirect. (p. 183)

This definition of a direct measure is problematic from a psychometric perspective because a direct self-assessment of a construct is never possible given that multiple items (questions) are^[16] by definition needed to measure a construct. Therefore, criterion 2 can never be respected apart from the trivial, and psychometrically deficient, case of using a single question to measure a construct¹. Using this definition virtually no measure in psychology can be classified as direct from a psychometric perspective and the distinction put forward by De Houwer and Moors (2010) would be of little utility. We think that the taxonomic distinction by De Houwer and Moors (2010) is very important but, to increase its usefulness, we propose to modify the definition of a direct measure. We define a direct measure as *a measurement procedure that is characterized by (a) a personal evaluation (e.g., questions such as “do you start conversations?” or “do you like chocolate?” requiring answers such as “very often” or “very much”) that is targeted to (b) an attribute (c) that could be included in the definition of the construct that the measurement outcome is assumed to capture (e.g., extraversion, attitude toward chocolate).*

The first property (*personal evaluation*) helps to differentiate a direct measure from a measure such as the IAT. The third property (*could be included in the definition of the construct*²) helps to differentiate standard questionnaires from measures such as the Name–Letter Task (NLT; Nuttin, 1985) that rely on a personal evaluation but that capture an attribute that would not be used to define the construct. In fact, starting conversations very often or affirming that one likes chocolate very much could be included in the definition of the constructs of extraversion and attitude toward chocolate, respectively. On the contrary, no one would include in the definition of self-esteem the preference for the letter of one’s name. In other words, the critical question here is to ask oneself whether one would use the measured outcome as a potential defining element of the construct: If the answer is no, the measure is indirect. Of course, often this is a continuum that we are dichotomizing only as a means to clarify the property. The second property (*an attribute*) helps to accommodate the fact that psychological measurement is generally characterized by two levels of abstraction, items and construct (e.g., Edwards & Bagozzi, 2000). Therefore, the measurement outcome is an element (an attribute) related to the construct rather than the construct itself (*the attribute*).

Using this definition as a benchmark, all measures should ideally have the second property (i.e., they are multi-items or stimuli), direct measures have all properties, whereas indirect measures do not have at least one among the first and the third properties. Moreover, this definition could be useful to further distinguish between different types of indirect measures depending on which of the two differentiating properties are missing. For instance, one could argue that the IAT does not have the first and the third property whereas the NLT has the_[17] first but not the third property. In fact, as we will detail later, a typical IAT is a task that is not characterized by a personal evaluation (e.g., it does not require one to express a personal opinion), similar to indirect measures such as the Affective Evaluative Priming (AEP; Fazio, Sanbonmatsu, Powell, & Kardes, 1986). The Affective Misattribution Paradigm (AMP; Payne, Cheng, Govorun, &

Stewart, 2005) and NLT instead rely on a personal evaluation (e.g., evaluate as positive or negative Chinese ideograms; evaluate alphabet letters) but the attributes they capture (e.g., preference for Chinese ideograms; preference for a letter) would not normally be used to define the construct (e.g., related to the primes in the AMP, self-esteem in the NLT).

What Is the Implicit Association Test?

The IAT is a paradigm that has been developed for the measurement of psychological constructs through the strength of associative links between concepts. It has been implemented to investigate a broad range of constructs (see the meta-analysis by Greenwald, Poehlman, Uhlmann, & Banaji, 2009). Unlike traditional interviews and questionnaires, in the IAT respondents are not requested to describe their own opinions or attitudes (e.g., by selecting their agreement to a question among several response options) but, rather, these are inferred based on their performance in a series of categorization tasks. Respondents see a series of stimuli appearing on a computer monitor (words or images) that represent two different (typically opposite) concepts and the two polarities of an attribute dimension. For each stimulus, they are required to press one of two different keys of the keyboard, depending on their category membership. For instance, in an IAT aimed to measure prejudice against Blacks, the two concepts could be the social categories Black and White, represented respectively by photographs of Black and White faces, and the attribute could be the positive–negative evaluation, represented by words (e.g., *rainbow*, *rotten*). The IAT is structured in different blocks. In the *simple categorization* blocks, the participants' task would be to press one key for White and the other for Black faces, or to press one key for positive and another key for negative words. Each stimulus belongs univocally to one category; the categorization task is therefore easy and the presence of an unambiguous relationship between each stimulus and its category is one of the prerequisites of a good implementation of IAT. The task is made more complex by the presence of