Evaluating the Construct Validity of Objective Personality Tests Using a Multitrait-Multimethod-Multioccasion-(MTMM-MO)-Approach

Article *in* European Journal of Psychological Assessment · January 2014

DOI: 10.1027/1015-5759/a000212

5 authors, including:

Tobias Koch
Friedrich Schiller University Jena

Tuulia M. Ortner
University of Salzburg

Manfred Schmitt
Universität Koblenz-Landau

Original Article

Evaluating the Construct Validity of Objective Personality Tests Using a Multitrait-Multimethod-Multioccasion-(MTMM-MO)-Approach

Tobias Koch,[1] T. M. Ortner,[2] M. Eid,[1] J. Caspers,[1] and M. Schmitt[3]

[1]Department of Education and Psychology, Freie Universität Berlin, Germany,
[2]Department of Psychology, Universität Salzburg, Austria,
[3]Department of Psychology, Universität Koblenz-Landau, Germany

Abstract. Although Objective Personality Tests (OPTs) have a long history in psychology and the field of psychological assessment, their validity, and reliability have not yet been sufficiently studied. In this study, we examined the convergent and discriminant validity of objective (personality) tests, Implicit Association Tests (IATs), and self-report measures for the assessment of conscientiousness and intelligence. Moreover, the convergent and discriminant validity of these measures was assessed on the trait (stable) and occasion specific (momentary) level by using the multimethod latent state-trait (MM-LST) model proposed by Courvoisier, Nussbeck, Eid, and Cole (2008) which allows for the decomposition of different sources of variance. Data from 367 students assessed on three different measurement occasions was incorporated. Results indicate generally low convergence of OPTs with data gained by other approaches. Additional analyses revealed that the OPTs used assess stable rather than momentary components of the constructs. Reliabilities of different tests ranged from .54 to .95. Furthermore, a substantial amount of trait method specificity revealed that different methods assess trait components that are not shared between OPTs and other measures. Data on the criterion validity of the objective conscientiousness test revealed that it is related to the punctuality of test takers in the laboratory.

Keywords: OPT, Objective Test, Objective Personality Test, MTMM, intelligence, conscientiousness

Objective personality tests (OPTs), defined as tests that deduce characteristics related to personality from observable behavior in performance tasks or other highly standardized miniature situations (Cattell & Warburton, 1967), have a long history in psychological measurement. They can even be traced back to the assessment pioneer, James McKeen Cattell, and his 10 tests based on "experiment and measurement" in 1890.

More recent, the history of OPTs is usually divided into two periods (see Ortner & Proyer, 2014): Starting in the 1940s, Raymond B. Cattell proposed to use so-called T-data for personality research. Within his theoretical framework, Cattell described objective (personality) tests as one of three sources of information within personality assessment besides self-report (Q-data) and biographic data (L-data; Cattell, 1946; Cattell & Kline, 1977). In line with this work, he created the *first generation* of OPTs, also known as *Cursive Miniature Situations* (Cattell, 1941, 1944) that stimulate behavioral expression of personality while fulfilling common standards of psychological tests. The proposed tests that also included standardized ability and achievement tests were characterized by a great variation with respect to material and scoring methods (Cattell, 1968).

During the last two decades, additional behavior-based methods of personality assessment, beyond self-report, have been developed, supported by the advances in computerized technologies. These tests, considered as the *second* generation of objective personality tests, have mostly benefitted from the development of computer technologies that provide highly flexible methods of item presentation and a precise registration of a person's behavior (Ortner et al., 2007; Ortner & Proyer, 2014). Examples of this are the inclusion of tachistoscopically presented stimuli (Proyer, 2007; Proyer & Häusler, 2007), as well as the assessment of viewing times, reaction times (e.g., Proyer, 2007), reaction speed (Schmidt-Atzert, 2007), animated consequences of behavior (Lejuez et al., 2002), or other

DOI: 10.1027/1015-5759/a000212

behaviors related to test performance such as guiding a figure in a maze (Ortner, Kubinger, Schrott, Radinger, & Litzenberger, 2006).

Several advantages regarding OPTs have been enumerated: Cattell (1968) suggested OPTs to assess information beyond self-report by fulfilling a "higher degree of objectivity." He proposed OPTs to measure how a person reacts to a standard stimulus situation "in ways of which he may not be aware and with interpretations of which he will certainly not be cognizant" (p. 165). Cattell also claimed that test takers cannot know themselves objectively enough to give a true picture (Cattell, 1968), as is typically required in personality questionnaires asking for self-reported behavior or attitudes. A common feature of today's OPTs is the principle to infer persons' individual characteristics from their overt behavior in such a standardized setting that lack face validity (see Cattell, 1968; Schmidt, 1975). Moreover, in case that personality is assessed in a narrower sense, masked aims of assessed variables of OPTs in fact turned out to establish lower susceptibility to manipulation compared to self-report measures, including faking as well as effects based on self-deception (Elliot, Lawty-Jones, & Jackson, 1996; Ziegler, Schmidt-Atzert, Bühner, & Krumm, 2007). This was obtained by the application of achievement-oriented tasks or particular scoring methods (see Cattell & Kline, 1977).

However, especially with regard to the second generations OPTs' validity, systematic, or large-scale approaches going beyond certain measures and single constructs are missing. OPTs did not even approximately gain the attention in research that, for example, another set of so-called *indirect* methods received, the best-known of these methods being the *Implicit Association Tests* (IAT; Greenwald, McGhee, & Schwartz, 1998). IATs have attracted considerable interest especially in social psychological research within the last years (e.g., Fazio & Olson, 2003; Rothermund & Wentura, 2004). One reason may be the large heterogeneity of tests in the OPT group with reference to, for example, task concepts, scoring, and materials. The present study aims to increase knowledge on newer OPTs' psychometric properties.

Prior Research on the Validity of OPTs

With respect to OPTs' criterion validity, data gained by several contemporary OPTs of the second generation indicate the utility of OPTs in research and practice in order to predict behavior-related indicators: As examples, in the domain of risk propensity, the Balloon Analog Risk Task (BART; Lejuez et al., 2002) is positively associated with risk-related behaviors such as smoking, gambling, drug and alcohol consumption, and risky sexual behaviors (Lejuez et al., 2002, 2003). The Crossing the Street Test (CtST; Santacreu, Rubio, & Hernández, 2006) predicts guessing tendency in a multiple-choice test in candidates taking an aptitude test (Rubio, Hernández, Zaldivar, Marquez, & Santacreu, 2010). More risky choices in patients with attention deficit hyperactivity disorder compared to a control group (Matthies, Philipsen, & Svaldi, 2012) were predicted by the Game of Dice Test (GDT; Brand et al., 2005). In the domain of achievement motivation, correlations with students' intermediate examination grade point average and school leaving examination grade point average served as indicators for the Objective Achievement Motivation Tests' predictive validity (OAT, Subtest 1; Schmidt-Atzert, 2004). Although these are only examples, results may suggest that criterion validity indicated by the existing relationship between OPTs and behavioral criteria seem to be successful at least for certain tests. Nevertheless, as OPTs considerably differ in test design, scoring procedure, and materials, results on validity regarding one test do not imply validity of others.

Findings on the construct validity of OPTs are mixed. A test's construct validity is given when empirical data confirm claims that were made based on a theory describing the given construct (Cronbach & Meehl, 1955). OPTs of the first generation were a result of Cattell's early theoretical proposal based on the idea that a complete description of personality requires the combination of Q-, L-, and T-data. Cattell originally expected to find that all three sources would point to a common underlying structure of personality represented by so-called source traits (Cattell, 1957). However, results gained by his decades-lasting research did not support his assumptions: The T-data and Q-data sets showed only low convergence. It was concluded that different methods assess different aspects of the underlying traits and that objective personality tests and questionnaires assessing self-report thus may systematically assess different aspects of personality, called, the "method-trait problem" (Hundleby, Pawlik, & Cattell, 1965). Similar results of nonconvergence including this first generation of tests were replicated by other early authors (Häcker, Schmidt, Schwenkmezger, & Utz, 1975; Häcker, Schwenkmezger, & Utz, 1979; Skinner & Howarth, 1973). Overall, OPTs showed little construct validity according to Cattell's own theory.

Lack of convergence or low convergence with self-report measures is a finding that was replicated for OPTs of the second generation: For example, studies using the Objective Achievement Motivation Test (Schmidt-Atzert, 2004) revealed zero correlations with the achievement striving scale of the NEO-PI-R (r = .02; Ziegler, Schmukle, Egloff, & Bühner, 2010). Dislich, Zinkernagel, Ortner, and Schmitt (2010) revealed lacking convergence in the domain of risk propensity between self-report data and the Balloon Analog Risk Task (BART; $r = -.17$; Lejuez et al., 2002), but some convergence between self-reported frequency of use of rationale calculation strategies and less risky choices in the Game of Dice task (r = .45; Brand, Heinze, Labudda, & Markowitsch, 2008).

The results of studies that investigated the validity of OPTs mirror closely findings that were obtained with indirect measures such as the IAT. First, a meta-analysis of the convergence between self-report scales and IAT measures of the same construct revealed a rather low average correlation of $r = .26$ (Hofmann, Gawronski, Gschwendner, Le, & Schmitt, 2005). Second, and in accordance with results that were obtained with OPTs, IATs seem to have incremental validity because they predict criterion variance that cannot be predicted with self-report scales (Friese, Hofmann, & Schmitt, 2008; Greenwald, Poehlman, Uhlmann, & Banaji, 2009). Third, and again in line with results that were obtained for OPTs, the convergence of different indirect measures tends to be low (e.g., Back et al., 2009; Bosson, Brown, Zeigler-Hill, & Swann, 2003; Dislich et al., 2010; Ziegler et al., 2010).

But what are potential causes for these results? One possible explanation for the low convergence of OPTs with other personality measures (e.g., self-report scales) may be that the given measures are confounded by unsystematic measurement error influences. Even in cases in which the measurement error has been taken into account, the low convergence among the measures may be due systematic method influences. Specifically, different measures may assess different aspects of the same construct (e.g., conscientiousness). Another possible explanation for the low convergence of OPTs and other measures may be due to the fact that the different measures assess stable as well as momentary aspects of a given construct. In fact, a varying amount of stability seems to be one of several reasons for the limited convergence between the IAT and self-report measures. For example, self-report scales tend to have high retest correlations that typically exceed .70, whereas much lower retest correlations have been found for the IAT despite its high internal consistency (Egloff, Schwerdtfeger, & Schmukle, 2005; Gschwendner, Hofmann, & Schmitt, 2008; Hofmann et al., 2005). This pattern suggests that the IAT does not only measure stable interindividual differences but also occasion specific (momentary) interindividual differences. The same may be true for OPTs.

So far, there has been no study examining the convergence of OPTs, indirect measures, and self-report scales across time while simultaneously accounting for stable and momentary method effects as well as unsystematic measurement influences. Moreover, only few studies have looked at the simultaneous convergence among OPTs, indirect measures, and self-report scales at all (Dislich et al., 2010).

The present study was designed to fill this gap using confirmatory factor analysis for longitudinal multitrait-multimethod (MTMM) data (see Courvoisier, Nussbeck, Eid, & Cole, 2008). The major advantage of this model is that is combines confirmatory factor analysis for MTMM data (CFA-MTMM models; see Eid, 2000; Eid, Lischetzke, Nussbeck, & Trierweiler, 2003; Eid, Nussbeck, Geiser, Cole, Gollwitzer, & Lischetzke, 2008) with latent state-trait (LST) theory (see Steyer, Schmitt, & Eid, 1999).

Structural Equation Modeling of Longitudinal MTMM Data

In the present study, the MTMM framework and the LST framework were combined in order to provide a more comprehensive validity analysis of the methods we chose to compare. LST is based on the fundamental assumption that psychological measure does not only assess stable individual differences (i.e., trait[1]), but also systematic individual differences that are due to the particular situation as well as the interaction between the person and the situation (see Steyer et al., 1999). As a generalization of the classical test theory, LST theory regards two decompositions: First, any observed variable (Y_{il}, where i = indicator and l = occasion of measurement) can be separated into a latent state (S_{il}) variable and a measurement error (ε_{il}) variable (Steyer et al., 1999). Second, any latent state (S_{il}) can be further decomposed into a latent trait (ξ_{il}) variable and a latent state-residual (ζ_{il}) variable representing situational effects and/or interaction effects (Steyer et al., 1999). According to LST theory, the total decomposition of any observed score is given by[2]:

$$Y_{il} = S_{il} + \varepsilon_{il}, \tag{1}$$

$$S_{il} = \xi_{il} + \zeta_{il}, \tag{2}$$

$$Y_{il} = \xi_{il} + \zeta_{il} + \varepsilon_{il}. \tag{3}$$

If the OPT assesses a trait, we would expect the results to be more or less stable and independent of the situation. If OPTs do not assess stable traits, this would necessarily lead to low convergence with trait measures. It has already been shown for IATs assessing extraversion and anxiety that they represent a higher amount of occasion specific variance than direct measures (Schmukle & Egloff, 2005). It is possible that OPTs also assess more occasion specific components than stable trait specific components. This would suggest that measures may then represent valid measures for the assessment of variable states, but not valid measures for the assessment of stable traits.

As questions of construct validity may combine questions of multiple methods and traits as well as a longitudinal perspective, Courvoisier et al. (2008) proposed a new structural equation model for MTMM data (called MM-LST model) combining the advantages of latent state-trait (LST) theory and the advantages of the correlated trait correlated method minus one [CTC ($M-1$)] modeling

[1] Throughout the entire article, the term ''trait'' is used to refer to stable (consistent) interindividual differences. The term ''construct'' is used to refer to particular psychological attributes or phenomena (e.g., conscientiousness or intelligence).

[2] According to LST theory, each latent state variable is decomposed into an indicator-specific and occasion specific latent trait and latent state-residual variable, respectively (see Steyer et al., 1999, p. 393).

approach (Eid, 2000; Eid et al., 2003, 2008). We choose the CTC ($M-1$) approach in this application, because it offers many advantages over other MTMM modeling approaches such as the latent difference (Lischetzke, Eid, & Nussbeck, 2002; Pohl, Steyer, & Kraus, 2008) or latent means approach (Pohl & Steyer, 2010). First, the CTC($M-1$) approach allows separating different variance components from one another and studying coefficients of true consistency and method specificity. Second, the CTC($M-1$) approach does not require that the different methods do not differ in their metric, which is of particular importance in this study given that different measures (i.e., self-reports, implicit measures, and OPTs) are used. For a detailed discussion on different modeling approaches of method effects in structural equation models see Geiser, Eid, West, Lischetzke, and Nussbeck (2012). A multimethod latent state-trait (MM-LST) model with two indicators, four methods, two constructs, and three occasions of measurements is depicted in Figure A1 of the Appendix.

According to the CTC($M-1$) approach, one method has to be chosen as reference (standard, comparison) method (see, e.g., Eid, 2000). In our example (see Figure A1), the first method ($k = 1$) has been selected as reference method. All other methods are then compared against this reference method. Within the MM-LST model (see Figure A1), the indicator-specific latent trait factors (T_{11}–T_{22}) represent stable trait effects as measured by the reference method (e.g., objective test). The occasion specific latent factors (O_{111}–O_{213}) reflect the momentary deviations from the stable latent trait factors as measured by the reference method. That is, the occasion specific latent factors compromise situational effects as well as interaction effects between the person and the situation as measured by the reference method. The indicator-specific latent trait *method* factors (TM_{112}–TM_{224},) pertaining to the nonreference method (e.g., self-reports) capture the consistent (stable) part of a nonreference indicator that cannot be predicted by the stable latent trait factors as measured by the reference method. Hence, the indicator-specific latent trait *method* factors reflect the consistent (general) deviation of the nonreference method (e.g., self-reports) from the stable latent trait factors as measured by reference method (e.g., objective test). The occasion specific latent *method* factors (OM_{121}–OM_{243}) capture the momentary deviations from the latent trait factors (occasion specific residuals) as measured by the nonreference method that cannot be predicted by the occasion specific latent factors as measured by the reference method. Hence, the occasion specific latent *method* factors reflect the momentary (occasion specific) over- or underestimation of the nonreference method with respect to the reference method. The complete measurement equation of the latent state variables (i.e., occasion-specific true score variables) can be expressed as follows:

$$S_{ijkl} = \begin{cases} T_{ij} + \lambda_{Oij1l} O_{j1l}, \quad \text{for } k = 1, \\ \alpha_{ijkl} + \lambda_{Tijkl} T_{ij} + \lambda_{Oijkl} O_{j1l} + \lambda_{TMijkl} \mathrm{TM}_{ijk} \\ \quad + \lambda_{OMijkl} \mathrm{OM}_{jkl}, \text{ for } k \neq 1, \end{cases} \tag{4}$$

where i = indicator, j = construct, k = method, and l = occasion of measurement. Again, note that with regard to the reference method ($k = 1$), neither latent trait (TM_{ijk}) nor occasion specific latent method factors (OM_{jkl}) are specified. These latent method factors are only defined for the nonreference methods in the CTC($M-1$) approach (Eid, 2000).

Due to the definition of the model, it is possible to calculate different variance components (see Courvoisier et al., 2008, pp. 274–275): the true (= measurement error free) trait consistency coefficient $\mathrm{TCon}(S_{ijkl})$, the true occasion specificity coefficient $\mathrm{OSpe}(S_{ijkl})$, the true trait method specificity coefficient $\mathrm{TMSpe}(S_{ijkl})$, the true occasion method specificity coefficient $\mathrm{OMSpe}(S_{ijkl})$, and the reliability coefficient $\mathrm{Rel}(Y_{ijkl})$. In Table 1 the formulas and definitions of all latent variables of the MM-LST model as well as the variance coefficients are provided.

In addition to the investigation of these different variance components, researchers may also study the discriminant validity on the trait as well as on the occasion specific (momentary) level. For example, high discriminant validity on trait level is indicated by low correlations among latent trait factors assessing different constructs (construct 1 = conscientiousness, construct 2 = intelligence). High discriminant validity on the occasion specific (momentary) level is reflected by low correlations among latent occasion specific factors assessing different constructs on the same occasion of measurement. In addition, researchers may also investigate the generalizability of method effects on the trait as well as the occasion specific (momentary) level by studying the correlations among the latent method factors on the trait level as well as on the occasion specific level. For a detailed description of all (non)admissible correlations as well as the conditions of parameter identification in the MM-LST model see Courvoisier et al. (2008).

In sum, the MM-LST model allows researchers to

a) evaluate the "true" (= measurement-error free) convergent and discriminant validity on the state- and trait-level among different methods,
b) investigate the extent to which particular constructs are "trait-like" versus "state-like,"
c) study the degree of stable as well as momentary method effects that are not shared with the reference method,
d) scrutinize whether or not stable or momentary method effects generalize across different constructs and methods,
e) relate other external variables (e.g., behavioral criteria) to the latent factors in the model.

In the present study, we used the MM-LST approach as proposed by Courvoisier et al. (2008) to investigate the construct validity (i.e., convergent and discriminant validity) of objective personality tests over time. In addition to OPTs, methods included self-report and indirect measures.

Table 1. Variance coefficients of the MM-LST model and their meaning (Courvoisier, 2006, pp. 82–130)

Formal definition	Description
$S_{ijkl} = \tau_{ijkl} = E(Y_{ijkl} \mid p_U, p_{Sit_l})$	Latent state variables (occasion specific true score variables). The latent state variables are defined as conditional expectations of the observed score given the person and the situation. $E(\cdot\mid\cdot)$ is the conditional expectation.
$\epsilon_{ijkl} = Y_{ijkl} - S_{ijkl} = Y_{ijkl} - E(Y_{ijkl} \mid p_U, p_{Sit_l})$	Measurement error variables are defined as difference between the observed scores and the latent state variables. By definition, these residual variables are uncorrelated with the true scores and have an expectation (mean) of zero.
$T_{ijkl} = E[(Y_{ijkl} \mid p_U, p_{Sit_l}) \mid p_U] = E(S_{ijkl} \mid p_U)$	Latent trait variables are defined as conditional expectations of the latent state variables given the person. Hence, the latent trait variables are free of occasional effects and only reflect person-specific (stable) effects.
$O_{ijkl} = S_{ijkl} - T_{ijkl} = E(Y_{ijkl} \mid p_U, p_{Sit_l}) - E(S_{ijkl} \mid p_U)$	Latent state-residual variables are defined as difference between the latent state variables and the latent trait variables. The latent state-residual variables capture the occasion specific effects as well as interaction effects between the person and the situation.
$E(T_{ijkl} \mid T_{ij1l}) = \alpha_{ijkl} + \lambda_{Tijkl} T_{ij}$	Latent regression on trait level. The latent trait variables pertaining to the nonreference method (e.g., self-reports) are regressed on the latent trait variables pertaining to the reference method (e.g., OPT).
$E(O_{ijkl} \mid O_{ij1l}) = \lambda_{Oijkl} O_{j1l}$	Latent regression on momentary level. The latent state-residual variables pertaining to the nonreference method (e.g., self-reports) are regressed on the latent state-residual variables pertaining to the reference method (e.g., OPT).
$\mathrm{TM}_{ijkl} = T_{ijkl} - E(T_{ijkl} \mid T_{ij1l}) = T_{ijkl} - \alpha_{ijkl} + \lambda_{Tijkl} T_{ij}$	Latent trait method variables are defined as part of the latent trait variables pertaining to the nonreference method that cannot be explained with respect to the latent trait variables pertaining to the reference method.
$\mathrm{OM}_{ijkl} = O_{ijkl} - E(O_{ijkl} \mid O_{ij1l}) = O_{ijkl} - \lambda_{Oijkl} O_{j1l}$	Latent state-residual method variables are defined as part of the latent state-residual variables pertaining to the nonreference method that cannot be explained with respect to the latent state-residual variables pertaining to the reference method.
$\mathrm{TM}_{ijkl} = \lambda_{\mathrm{TM}ijkl} \mathrm{TM}_{jkl}$	Homogeneity assumption of the latent trait method variables. It is assumed that the latent trait method variables are positive linear transformations of each other and only differ by a multiplicative constant (λ_{TMijkl}).
$\mathrm{OM}_{ijkl} = \lambda_{\mathrm{OM}ijkl} \mathrm{TM}_{jkl}$	Homogeneity assumption of the latent state-residual method variables. It is assumed that the latent state-residual method variables are positive linear transformations of each other and only differ by a multiplicative constant (λ_{OMijkl}).
$\mathrm{Var}(Y_{ij1l}) = \lambda^2_{Tij1l} \mathrm{Var}(T_{ij}) + \lambda^2_{Oijkl} \mathrm{Var}(O_{j1l}) + \mathrm{Var}(\epsilon_{ij1l})$	Variance Decomposition (of the reference method indicators).
$\mathrm{Var}(Y_{ijkl}) = \lambda^2_{Tijkl} \mathrm{Var}(T_{ij1l}) + \lambda^2_{Oijkl} \mathrm{Var}(O_{ij1l}) + \lambda^2_{\mathrm{TM}ijkl} \mathrm{Var}(\mathrm{TM}_{ijkl}) + \lambda^2_{\mathrm{OM}ijkl} \mathrm{Var}(\mathrm{OM}_{ijkl}) + \mathrm{Var}(\epsilon_{ijkl})$	Variance Decomposition (of the nonreference method indicators).

(Continued on next page)

Table 1. (Continued)

Formal definition	Description
$\mathrm{TCon}(S_{ijkl}) = \frac{\lambda^2_{Tijkl}\mathrm{Var}(T_{ij})}{\mathrm{Var}(Y_{ijkl})-\mathrm{Var}(\varepsilon_{ijkl})}$	Proportion of true variance due to true stable (consistent) interindividual differences measured by the reference method.
$\mathrm{OSpe}(S_{ijkl}) = \frac{\lambda^2_{Oijkl}\mathrm{Var}(O_{j1l})}{\mathrm{Var}(Y_{ijkl})-\mathrm{Var}(\varepsilon_{ijkl})}$	Proportion of true variance due to true momentary (occasion specific) interindividual differences measured by the reference method.
$\mathrm{TMSpe}(S_{ijkl}) = \frac{\lambda^2_{TMijkl}\mathrm{Var}(\mathrm{TM}_{ijk})}{\mathrm{Var}(Y_{ijkl})-\mathrm{Var}(\varepsilon_{ijkl})}$	Proportion of true variance due to true stable (consistent) interindividual differences measured by a nonreference method indicator that cannot be predicted by the latent trait factor of the reference method.
$\mathrm{OMSpe}(S_{ijkl}) = \frac{\lambda^2_{OMijkl}\mathrm{OM}_{jkl}}{\mathrm{Var}(Y_{ijkl})-\mathrm{Var}(\varepsilon_{ijkl})}$	Proportion of true variance due to true momentary (occasion specific) interindividual differences measured by a nonreference method indicator that cannot be predicted by the occasion specific latent factors of the reference method.
$\mathrm{Rel}(Y_{ijkl}) = 1 - \frac{\mathrm{Var}(E_{ijkl})}{\mathrm{Var}(Y_{ijkl})} = \frac{\mathrm{Var}(\tau_{ijkl})}{\mathrm{Var}(Y_{ijkl})}$	Proportion of observed variance due to true variance.

Aims of the Present Study

Systematic large-scale approaches investigating the validity of OPTs, such as those that were conducted during their early development period, have not been undertaken in the past few decades using longitudinal MTMM structural equation models. The following study aims to contribute to a better understanding of what OPTs measures (a) by scrutinizing the convergent and discriminant validity and reliability of OPTs and other personality measures across time, (b) by evaluating the amount of stable and momentary influences of OPT measures, and (c) by assessing aspects of the criterion validity of the OPT measures with respect to real-life criteria. We seek to answer these questions more comprehensively as we combine the "classical" *multi-trait-multimethod* (MTMM; Campbell & Fiske, 1959) approach with structural equation modeling (SEM) and longitudinal modeling approaches (here: latent state-trait theory). We investigate the psychometric properties of the OPT measures over time by using the multimethod latent state-trait (MM-LST) model proposed by Courvoisier et al. (2008).

In our study, we focused on the assessment of two constructs, conscientiousness and general intelligence, with the former representing personality in a narrower sense and the latter in a wider range of interpretation. Conscientiousness has been defined as the tendency to show self-discipline, act dutifully, to aim for achievement, and to show a preference for planned rather than spontaneous behavior (Costa & McCrae, 1992). General intelligence has been defined as the ability to act thoughtfully and in an intelligent way, and furthermore, to draw inferences from given premises to provide a conclusion (Carroll, 1993). Whereas the first construct is typically assessed via self-report in psychological research and practice, the latter is normally assessed by standardized tests. Within our research design we aimed to investigate the convergence of OPTs with two other groups of measures: Each construct was therefore assessed by a total of four methods: two OPT measures, one indirect measure (IAT), and one direct measure (self-report questionnaire). We further aimed for a tight conceptual correspondence of the assessed construct across the different methods. The assessment was repeated on three occasions of measurement.

The data analyses were performed in three steps: First, taking measurement error into account, we examined the convergent validity of the OPTs with each other, and with indirect and direct measures for each construct (conscientiousness and general intelligence) separately. This was done by specifying monoconstruct MM-LST models. Along with these models, we gained estimates of the method specificity on each measurement occasions (i.e., occasion specific method specificity, OMSpe) and across measurement occasions (i.e., trait method specificity, TMSpe). Second, we evaluated the discriminant validity as well as the generalization of method effects across constructs by specifying multiconstruct MM-LST models for different combinations of methods. There were two reasons for using multiple reduced MM-LST models for answering our research questions rather than using one large MM-LST model: First, the MM-LST submodels imply fewer parameters to be estimated for the same number of observations. Second, the parameter estimates remain unaffected if the submodels of a larger MM-LST model are specified and the same reference method is chosen (see Geiser et al., 2012). A general recommendation in structural equation modeling is to sample at least five observations per model parameter in order to obtain reliable parameter estimates (Bentler & Chou, 1987). This general rule of thumb is more likely to be met by specifying reduced MM-LST models (ratio in this study 5.31 for conscientiousness and 4.89 for intelligence). Finally, the criterion validity of the OPTs measures for the assessment of conscientiousness was tested by including a variable measuring punctuality into the model. To our knowledge, the extension of the MM-LST model with regard to the inclusion of external real life criteria has not been proposed so far.

Materials and Methods

Overview

Data was collected at two study sites in Austria and Germany from 2011 to 2013. Participants were tested three times: The second time took place a minimum of 1 week and a maximum of 2 weeks after the first test session, and the third time again took place a minimum of 1 week and a maximum of 2 weeks after the second test session. Test sessions were conducted in the computer laboratory in six workplaces at the two universities. Textile partitions were used to separate the workplaces. After the last session, participants were debriefed and were offered the possibility of receiving personal information about their results.

Participants

Three-hundred-sixty-seven university students (278 female, 89 male) aged 16–36 years (M = 21.93, SD = 3.91) participated in 2011 and 2012 in exchange for a certificate of research participation and/or monetary compensation.

Materials

Conscientiousness

Objective Personality Tests

The first test employed for the assessment of conscientiousness was the COPT (Conscientiousness Objective Personality Test) based on a procedure proposed by Steffens and König (2006). Test takers were instructed to work on an excerpt from a common attention test (Brickenkamp, 2002) that required crossing out each *d* accompanied by two primes among similar-looking letters with various numbers of primes, in total, 329 stimuli. In contrast to the original attention test which works with time constraint, test takers were only instructed to work "as accurately as possible." A lower number of mistakes was considered to be an indicator of conscientiousness.

The second test, called the C62 (Conscientiousness test 62), was based on a modified version of the T62 test "Hesitancy" by Cattell and Warburton (1967). In this test, test takers view pairs of figures on a screen and are instructed to select the larger one; after a decision has been made, a new pair of figures appears on the screen (see also Kubinger & Ebenhöh, 1996). The figures only differ by about 10% of their total size but are rotated to different angles. Again, test takers were instructed to make accurate decisions. A total number of 20 items was employed and time spent per item served as scores.

Implicit Association Test

We included a single category IAT (SC-IAT) in order to assess the implicit self-concept of conscientiousness (CIAT). SC-IATs are characterized by only a single target category rather than two target categories (SC-IAT; Karpinski & Steinman, 2006). We employed stimuli proposed by Schmukle, Back, and Egloff (2008) that were extracted from a list of 823 adjectives assigned to the Big Five personality dimensions (Ostendorf, 1994). We included "me" as target category and "conscientious" versus "careless" as attribute categories. Words such as "accurate," "careful," and "tidy" represented the attribute category conscientious; "careless" was represented by words such as "faithless," "chaotic," and "messy."[3] The SCI-IAT was scored according to the improved scoring algorithm (D600) proposed by Greenwald, Nosek, and Banaji (2003). Combined blocks consisted of 75 trials. High (low) SC-IAT scores indicate strong associations between "me" and "carelessness" ("conscientious") and thus indicate an implicit careless (conscientious) self-concept.

Self-Report

Ten items (e.g., *I think thoroughly through things before making a decision*; *I try to complete task very accurately, so that they do not have to be completed again*) were taken from the NEO-PI-R (Ostendorf & Angleitner, 2003). Responses were recorded on 5-point rating scales ranging from 1 (= *not at all*) to 5 (= *completely*).

Criterion

We measured students' average punctuality at test session 2 and session 3 as behavioral criterion. Time of arrival was

[3] All original stimuli used in this study were presented in the German language.

registered in minutes, with the number of minutes arriving early coded negatively (e.g., arrival 3 min too early = −3), and the number of minutes arriving late coded positively.

Intelligence

Objective Test

As the first test we implemented the "Number Series" (NST) task of the German *Intelligenz-Struktur-Test 2000 R* (I-S-T 2000 R; Amthauer, Brocke, Liepmann, & Beauducel, 2001) consisting of 20 items. This test aims to assess numeric reasoning. Test takers were instructed to identify and write down the next number corresponding to a rule. Time to complete the test was restricted to 10 min.

As a second test, a fixed item short version of the *Adaptive Matrices Test* (AMT; Hornke, Küppers, & Etzel, 2000) was employed. The AMT is a nonverbal computer-based test that assesses reasoning on the basis of figural items and it fulfills the criteria of the Rasch model (Rasch, 1960). The test was constructed on the basis of a theoretical rationale defining item generating rules constituting items' difficulties. The version we employed contained 16 items with increasing difficulty (item parameters were −3.58, −3.19, −2.57, −2.29, −1.71, −1.35, −0.86, −0.54, −0.37, −0.05, 0.22, 0.41, 0.57, 1.49, 1.56, 2.20). Again, time to complete the test was restricted to 10 min.

Implicit Association Test

We also included a single category IAT (SC-IAT) in order to assess the implicit self-concept of intelligence (IIAT). We employed stimuli proposed by Dislich et al. (2012). We included "me" as target category and "intelligent" versus "stupid" as attribute categories. Words such as "clever," "bright," and "able" represented the attribute category intelligent; "stupid" was represented by words such as "silly," "dumb," and "foolish." Combined blocks consisted of 75 trials. High (low) SC-IAT scores indicate strong associations between "me" and "stupid" ("intelligent").

Self-Report

All items of this set (SR-INT) were self-designed (e.g., *I do not consider myself as intelligent*; *I am good in finding a solution for a given complex problem*). Responses were recorded on a 5-point rating scales ranging from 1 (= *not at all*) to 5 (= *completely*).

Procedure

Testing was conducted by up to four administrators at each study site. General information was given about the duration and parts of the testing. Subsequently, an information sheet was read to the students. Furthermore, participants were randomly assigned to one of four groups differing in the order in which measures were presented. To prevent test takers from making inferences about our measurement aims, rating scales were presented after all OPTs and IATs had been finished.

Statistical Analysis

Factor Indicators

For the statistical analyses (i.e., MM-LST model) we used parcels instead of single items as indicators. There were two main reasons for using parcels instead of single items as indicators. First, some of the methods described above (e.g., NST, AMT) were measured by categorical (dichotomous) observed variables rather than continuous observed variables. However, complex structural equation models with categorical observed variables are computationally extremely demanding and require large sample sizes in order to obtain proper parameter estimates (e.g., Finney & DiStefano, 2006; Muthén & Muthén, 1998–2012; Nussbeck, Eid, & Lischetzke, 2006). Second, some of the measures (e.g., NST, AMT) described above consisted of 10–20 items, which would have led to extremely complex models. Therefore, we created two parcels (test halves) for each construct-method-occasion unit (CMOU). The parcels were built based on the results of extensive preliminary item analyses (cf. Little, Cunningham, Shahar, & Widaman, 2002).

Specification of the MM-LST Models

In the first step, we specified separate MM-LST models for each construct. Both models are depicted in Figures 1 and 2.

For the sake of clarity, the trait and occasion specific level of each model is illustrated on separate layers in Figures 1 and 2 (white layer = trait level; gray layer = occasion specific level). For each model, we specified indicator-specific latent trait factors (T_{11}, T_{21}, T_{12}, and T_{22}) as well as indicator-specific latent trait method factors (TM_{112}, TM_{212}, TM_{113}, TM_{213}, TM_{114}, TM_{214}, TM_{122}, TM_{222}, TM_{123}, and TM_{223}) on trait level (white layer), except for the latent trait method factor (TM_{24}) belonging to the self-report measures of intelligence. With regard to the implicit measures, it was possible to specify one common latent trait method factor rather than two indicator-specific latent trait method factors, given that the two indicator-specific latent trait method factors (TM_{123} and TM_{223}) were perfectly correlated with each other. The two objective tests, COPT and NST, were used as reference method, following the CTC($M - 1$) modeling approach. The remaining methods C62, CIAT, SR-NEO, AMT, IIAT, and SR-INT served as nonreference methods and were therefore contrasted against the reference method traits (see the right-hand side of Figures 1 and 2).

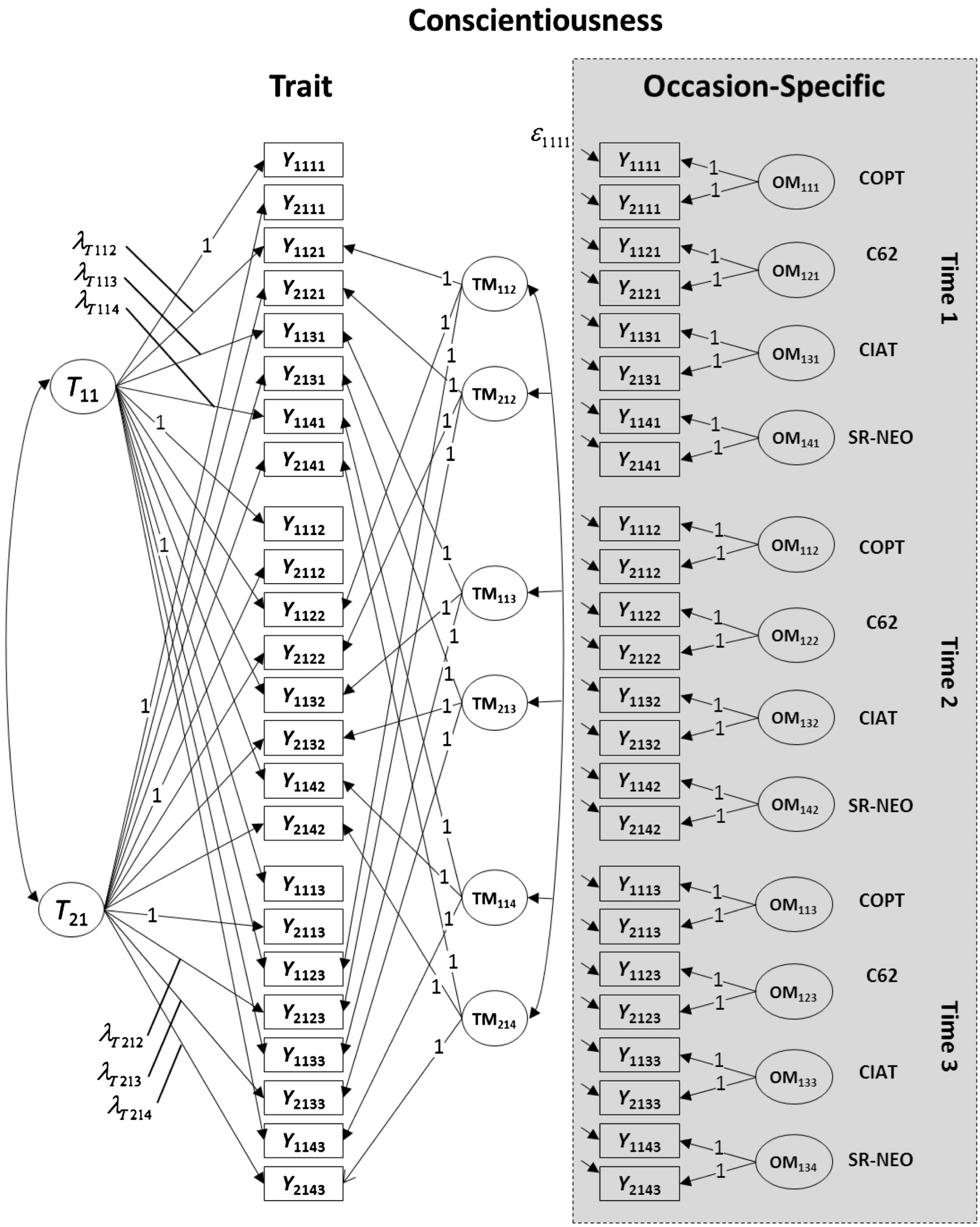

Figure 1. Multitrait-multimethod latent state-trait (MM-LST) model for conscientiousness measured by two indicators and four methods on three occasions of measurement. T_{ij} = reference trait factor; TM_{ijk} = latent trait method factors; OM_{jkl} = occasion specific method factors; ε_{ijkl} = measurement error variables; λ_{Tijk} = time-invariant factor loadings; i = indicator; j = construct; k = method; l = occasion of measurement. COPT = Conscientiousness Objective Personality Test; C62 = Conscientiousness-Test 62; CIAT = implicit self-concept of conscientiousness; SR-NEO = self-report questionnaire for conscientiousness.

On the occasion specific level, it was not feasible to specify a CTC($M-1$) structure. This was mainly due to extremely low momentary (occasion specific) influences (see Tables 4 and 5) and zero correlations between the latent occasion specific factors. Therefore, the convergent validity was zero on the occasion specific level.

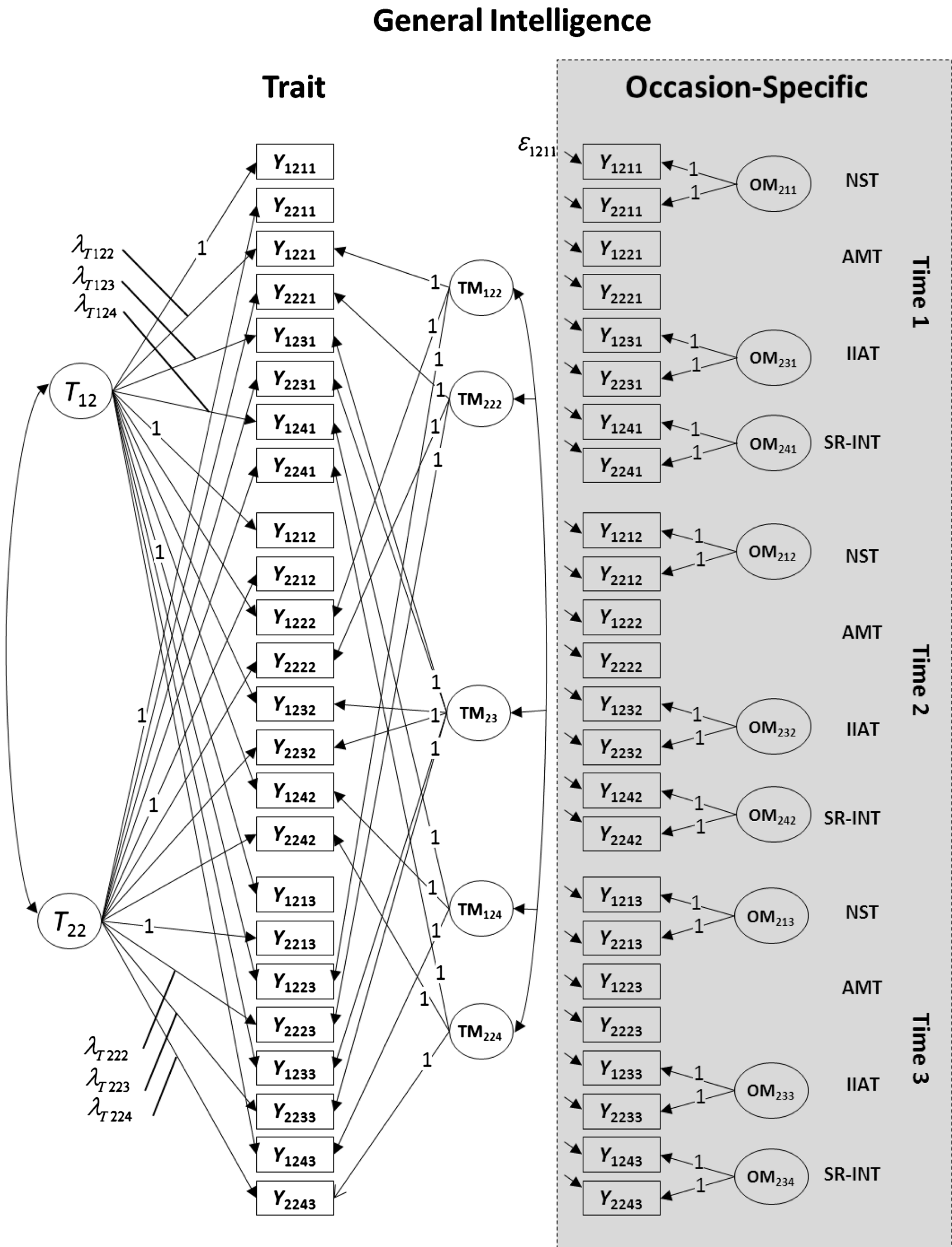

Figure 2. Multitrait-multimethod latent state-trait (MM-LST) model for general intelligence measured by two indicators and four methods on three occasions of measurement. T_{ij} = reference trait factor; TM_{ijk} = latent trait method factors; OM_{jkl} = occasion specific method factors; ε_{ijkl} = measurement error variables; λ_{Tijk} = time-invariant factor loadings; i = indicator; j = construct, k = method; l = occasion of measurement. NST = Number Series Test; AMT = Adaptive Matrices Test; IIAT = implicit self-concept of intelligence; SR-INT = self-report questionnaire for general intelligence. Due to the high stability of the AMT, the occasion specific latent factors have been fixed to zero in the analysis and thus are omitted in the figure.

Table 2. Unstandardized intercepts, factor loadings, and error variances for the MM-LST model (Conscientiousness)

Parcels (Test halves)	Intercepts (α_{ijkl})	Trait factor loading	Trait method factor loading	Occ.-specific factor loading	Error variance
		Conscientiousness			
Ref. Method 1 (COPT)					
Y_{1111}	0.71[a]	1.00[b]	–	1.00[b]	0.04[a]
Y_{2111}	0.73[a]	1.00[b]	–	1.00[b]	0.05[a]
Y_{1112}	0.81[a]	1.00[b]	–	1.00[b]	0.03[a]
Y_{2112}	0.80[a]	1.00[b]	–	1.00[b]	0.04[a]
Y_{1113}	0.84[a]	1.00[b]	–	1.00[b]	0.03[a]
Y_{2113}	0.82[a]	1.00[b]	–	1.00[b]	0.04[a]
Nonref. Method 2 (C62)					
Y_{1121}	0.73[b]	0.07[b]	1.00[b]	1.00[b]	0.04[a]
Y_{2121}	0.83[b]	0.19[b]	1.00[b]	1.00[b]	0.03[a]
Y_{1122}	0.73[b]	0.07[b]	1.00[b]	1.00[b]	0.04[a]
Y_{2122}	0.83[b]	0.19[b]	1.00[b]	1.00[b]	0.04[a]
Y_{1123}	0.73[b]	0.07[b]	1.00[b]	1.00[b]	0.04[a]
Y_{2123}	0.83[b]	0.19[b]	1.00[b]	1.00[b]	0.03[a]
Nonref. Method 3 (CIAT)					
Y_{1131}	0.25[b]	−0.18[b]	1.00[b]	1.00[b]	0.06[a]
Y_{2131}	0.25[b]	−0.23[b]	1.00[b]	1.00[b]	0.06[a]
Y_{1132}	0.25[b]	−0.18[b]	1.00[b]	1.00[b]	0.05[a]
Y_{2132}	0.25[b]	−0.23[b]	1.00[b]	1.00[b]	0.04[a]
Y_{1133}	0.25[b]	−0.18[b]	1.00[b]	1.00[b]	0.06[a]
Y_{2133}	0.25[b]	−0.23[b]	1.00[b]	1.00[b]	0.09[a]
Nonref. Method 4 (SR-NEO)					
Y_{1141}	3.65[b]	0.15[b]	1.00[b]	1.00[b]	0.07[a]
Y_{2141}	3.88[b]	−0.11[b]	1.00[b]	1.00[b]	0.05[a]
Y_{1142}	3.65[b]	0.15[b]	1.00[b]	1.00[b]	0.07[a]
Y_{2142}	3.88[b]	−0.11[b]	1.00[b]	1.00[b]	0.03[a]
Y_{1143}	3.65[b]	0.15[b]	1.00[b]	1.00[b]	0.06[a]
Y_{2143}	3.88[b]	−0.11[b]	1.00[b]	1.00[b]	0.04[a]

Notes. Y_{ijkl} = observed variable; i = indicator; j = construct; k = method; and l = occasion of measurement; [a] = parameter that is freely estimated across time; [b] = parameter that is set to be equal across time.

Consequently, uncorrelated occasion specific latent method factors were assumed. With regard to the AMT, the variance of the occasion specific factor approached zero. Therefore, we fixed this occasion specific factor to zero in the analysis.

In the second step, we analyzed the discriminant validity by combining both single construct MM-LST models. In the third step, we scrutinized the criterion validity by including behavioral outcome data (i.e., lateness on arrival in seconds) regarding the conscientiousness OPTs into the model.

Goodness-of-Fit Assessment

The goodness-of-fit of models was determined by taking the following fit criteria into consideration: the chi-square test, the root mean square error of approximation (RMSEA; Steiger, 1990), the comparative fit index (CFI; Bentler, 1990), and the standardized root mean square residual (Bentler, 1995). Models were said to fit the data, if the following criteria were met:

(a) Nonsignificant chi-square value or the ratio of chi-square to its degrees of freedom below two (see Schermelleh-Engel, Moosbrugger, & Müller, 2003).
(b) An RMSEA value equal to or less than .05 (Hu & Bentler, 1998, 1999).
(c) A CFI value equal to or greater than .97 (Hu & Bentler, 1998, 1999; Schermelleh-Engel et al., 2003).
(d) An SRMR (standardized root mean square residual) below .08 (Hu & Bentler, 1998, 1999).

Statistical Software and Estimators

The preliminary analyses (i.e., descriptive analyses, item analyses, parceling) were using the software R, Version 2.15.2 (R Development Core Team, 2013). The multimethod latent state-trait (MM-LST) models were specified and estimated in M*plus*, Version 7.0 (Muthén & Muthén, 1998–2012). All models were estimated using the robust maximum likelihood (MLR) estimator and full information maximum likelihood (FIML) estimation to account for missing data.

Table 3. Unstandardized intercepts, factor loadings, and error variances for the MM-LST model (General Intelligence)

Parcels (Test halves)	Intercepts (α_{ijkl})	Trait factor loading	Trait method factor loading	Occ.-specific factor loading	Error variance
		General Intelligence			
Ref. Method 1 (NST)					
Y_{1211}	0.79[a]	1.00[b]	–	1.00[b]	0.02[a]
Y_{2211}	0.79[a]	1.00[b]	–	1.00[b]	0.02[a]
Y_{1212}	0.87[a]	1.00[b]	–	1.00[b]	0.01[a]
Y_{2212}	0.88[a]	1.00[b]	–	1.00[b]	0.01[a]
Y_{1213}	0.90[a]	1.00[b]	–	1.00[b]	0.01[a]
Y_{2213}	0.89[a]	1.00[b]	–	1.00[b]	0.01[a]
Nonref. Method 2 (AMT)					
Y_{1221}	0.51[a]	0.31[b]	1.00[b]	–	0.02[a]
Y_{2221}	0.46[a]	0.49[b]	1.00[b]	–	0.02[a]
Y_{1222}	0.53[a]	0.31[b]	1.00[b]	–	0.01[a]
Y_{2222}	0.50[a]	0.49[b]	1.00[b]	–	0.02[a]
Y_{1223}	0.53[a]	0.31[b]	1.00[b]	–	0.02[a]
Y_{2223}	0.51[a]	0.49[b]	1.00[b]	–	0.02[a]
Nonref. Method 3 (IIAT)					
Y_{1231}	0.36[a]	−0.14[b]	1.00[b]	1.00[b]	0.05[a]
Y_{2231}	0.38[a]	−0.08[b]	1.00[b]	1.00[b]	0.07[a]
Y_{1232}	0.27[a]	−0.14[b]	1.00[b]	1.00[b]	0.04[a]
Y_{2232}	0.27[a]	−0.08[b]	1.00[b]	1.00[b]	0.04[a]
Y_{1233}	0.24[a]	−0.14[b]	1.00[b]	1.00[b]	0.04[a]
Y_{2233}	0.22[a]	−0.08[b]	1.00[b]	1.00[b]	0.04[a]
Nonref. Method 4 (SR-INT)					
Y_{1241}	3.19[a]	2.27[b]	1.00[b]	1.00[b]	0.11[a]
Y_{2241}	3.48[a]	3.00[b]	1.00[b]	1.00[b]	0.14[a]
Y_{1242}	3.30[a]	2.27[b]	1.00[b]	1.00[b]	0.06[a]
Y_{2242}	3.59[a]	3.00[b]	1.00[b]	1.00[b]	0.10[a]
Y_{1243}	3.39[a]	2.27[b]	1.00[b]	1.00[b]	0.05[a]
Y_{2243}	3.64[a]	3.00[b]	1.00[b]	1.00[b]	0.07[a]

Notes. Y_{ijkl} = observed variable; i = indicator; j = construct; k = method; and l = occasion of measurement; [a] = parameter that is freely estimated across time; [b] = parameter that is set to be equal across time.

Results

Convergence, Method Specificity, and Reliability

The monoconstruct MM-LST model for conscientiousness fit the data well, $\chi^2(255, N = 367) = 285.04$, $p = .09$, CFI = .99, RMSEA = .02, SRMR = .05. The single construct MM-LST model for intelligence fit the data acceptably well, $\chi^2(249, N = 367) = 351.96$, $p < .001$, CFI = .98, RMSEA = .03, SRMR = .07. The unstandardized intercepts, factor loadings, and latent error variances of the MM-LST models are provided in Tables 2 and 3. According to these tables, both models imply invariance restrictions with respect to the factor loadings and intercepts pertaining to the same indicator, construct, and method across time. In addition, the variances of the occasion specific latent method factors were set equal across all occasions of measurement. No measurement invariance restrictions were imposed with regard to the measurement error variances. Invariance restrictions regarding the intercepts of the indicators could not be established for all indicators (see Tables 2 and 3). Specifically, the intercepts pertaining to the indicators of the COPT, NST, AMT, IIAT, and SR-INT were freely estimated (not set to be invariant across time). This means that MM-LST allows for learning effects with regard to these measures. For a detailed discussion of measurement invariance conditions in LST models and their consequences see Geiser et al. (2014).

In Tables 4 and 5, the standardized factor loadings as well as the different variance coefficients are presented for the same models. For the sake of clarity, the results given in the Tables 2–5 will be explained step by step, thus linking the particular parameters to the path diagrams given in Figures 1 and 2. According to Figure 1, the conscientiousness indicators Y_{1111}, Y_{2111}, Y_{1112}, Y_{2112}, Y_{1113}, and Y_{2113} pertaining to the first method (i.e., COPT) were indicators of two indicator-specific latent trait factors (T11, T21) as well as three occasion specific latent method factors (OM_{111}, OM_{112}, OM_{113}). Due to the definition of the MM-LST model, the total variance of each of these indicators could be decomposed into three components: the trait consistency (TCon), the occasion method specificity (OMSpe), and the unreliability $[1 - \mathrm{Rel}(Y_{ijkl})]$. As indicated before, the occasion method specificity coefficients (OMSpe) reflect interindividual differences due to momentary influences. In contrast, the trait consistency coefficient (TCon)

Table 4. Standardized factor loading parameters, trait consistency, convergent correlations, trait method specificity, occasion specificity, and reliability for Conscientiousness

Parcels (Test halves)	Trait factor loading	Trait method factor loading	Occ.-specific factor loading	TCon	Corr	TMSpe	OMSpe	Rel	CRel
			Conscientiousness						
Ref. Method 1 (COPT)									
Y_{1111}	.63	–	.34	.78	–	–	.22	.52	
Y_{2111}	.62	–	.33	.78	–	–	.22	.50	.68
Y_{1112}	.70	–	.38	.77	–	–	.23	.63	
Y_{2112}	.66	–	.35	.78	–	–	.22	.55	.75
Y_{1113}	.71	–	.38	.77	–	–	.23	.65	
Y_{2113}	.67	–	.35	.78	–	–	.22	.57	.76
Nonref. Method 2 (C62)									
Y_{1121}	.05	.63	.21	.01	.10	.90	.09	.45	
Y_{2121}	.16	.63	.25	.06	.24	.81	.13	.49	.64
Y_{1122}	.05	.62	.20	.01	.10	.90	.09	.43	
Y_{2122}	.14	.55	.22	.06	.24	.82	.12	.37	.57
Y_{1123}	.05	.65	.21	.01	.10	.90	.09	.48	
Y_{2123}	.16	.61	.24	.06	.24	.81	.13	.46	.64
Nonref. Method 3 (CIAT)									
Y_{1131}	−.10	.48	.53	.02	.14	.44	.54	.51	
Y_{2131}	−.13	.42	.53	.04	.20	.37	.59	.48	.66
Y_{1132}	−.11	.50	.56	.02	.14	.44	.54	.57	
Y_{2132}	−.14	.45	.57	.03	.17	.38	.59	.55	.72
Y_{1133}	−.10	.46	.51	.02	.14	.44	.54	.48	
Y_{2133}	−.12	.38	.48	.04	.20	.37	.59	.39	.61
Nonref. Method 4 (SR-NEO)									
Y_{1141}	.04	.91	.19	.00	.00	.96	.04	.86	
Y_{2141}	−.04	.90	.22	.00	.00	.94	.06	.86	.92
Y_{1142}	.04	.91	.19	.00	.00	.96	.04	.86	
Y_{2142}	−.04	.92	.23	.00	.00	.94	.06	.90	.94
Y_{1143}	.04	.91	.19	.00	.00	.96	.04	.87	
Y_{2143}	−.04	.92	.22	.00	.00	.94	.06	.89	.94

Notes. TCon = true trait consistency; Corr = $\sqrt{TCon}$ (correlation between measures on trait level); TMSpe = true trait method specificity; OMSpe = true occasion method specificity; Rel = reliability; CRel = composite reliability. Y_{ijkl} = observed indicator (i.e., parcels); i = indicator; j = construct; k = method; and l = occasion of measurement. Note that the coefficient of consistency, method specificity may not add up to 1.00 in the above table due to rounding errors.

captures the proportion of "true" variance of an indicator that is due to "true" stable (consistent) interindividual differences (Courvoisier et al., 2008). With regard to the reference method indicators, the "true" trait consistency and the "true" occasion specificity coefficients add up to 1. For example, the "true" trait consistency coefficient of the first reference method indicator (Y_{1111}) was .78, the occasion specificity coefficient was .22 (see Table 4).

That means that 78% of the true variance of the first reference method indicator (Y_{1111}) was determined by stable (consistent) interindividual differences. In contrast, 22% of the true variance of this indicator was due to occasion specific (momentary) influences. The "true" trait consistency coefficient (TCon) can be obtained by the squared value of the standardized trait factor loading divided by the true variance in the standardized model (i.e., observed variance minus measurement error variance, see Table 1). Similarly, the occasion method specificity (OMSpe) coefficient is obtained by the squared value of the standardized occasion specific factor loading divided by the true variance in the standardized model. Specifically, the trait consistency coefficient for the first indicator Y_{1111} was given by: $(.63)^2/((.63)^2 + (.34)^2) = .78$. The occasion method specificity coefficient was computed analogously: $(.34)^2/((.63)^2 + (.34)^2) = .22$. According to these results, the trait consistency coefficients were always larger than the occasion method specificity coefficients for the reference method indicators (see Table 4). This indicates that the reference method indicators (i.e., COPT) assess stable (trait) rather than momentary (state-like) components of conscientiousness (see Table 4). In a similar way, the trait and occasion specificity coefficients could be obtained for the intelligence indicators pertaining to the reference method (i.e., NST).

For the first intelligence indicator (Y_{1211}), the "true" trait consistency coefficient equaled $(.67)^2/((.67)^2 + (.23)^2) = .90$, whereas the "true" occasion specificity coefficient was $(.23)^2/((.67)^2 + (.23)^2) = .10$. This means that 90% of true variance of the intelligence indicator Y_{1211} was due to stable (consistent) interindividual differences, whereas 10% of true variance was due to occasion specific (momentary) interindividual differences. The reliability

Table 5. Standardized factor loading parameters, trait consistency, convergent correlations, trait method specificity, occasion specificity, and reliability for General Intelligence

Parcels (Test halves)	Trait factor loading	Trait method factor loading	Occ.-specific factor loading	TCon	Corr	TMSpe	OMSpe	Rel	CRel
			General Intelligence						
Ref. Method 1 (NST)									
Y_{1211}	.67	–	.23	.90	–	–	.10	.50	
Y_{2211}	.71	–	.23	.90	–	–	.10	.56	.69
Y_{1212}	.81	–	.27	.90	–	–	.10	.74	
Y_{2212}	.81	–	.26	.91	–	–	.09	.72	.84
Y_{1213}	.85	–	.29	.90	–	–	.10	.81	
Y_{2213}	.80	–	.26	.90	–	–	.10	.71	.87
Nonref. Method 2 (AMT)									
Y_{1221}	.25	.48	–	.22	.47	.78	–	.29	
Y_{2221}	.33	.57	–	.25	.50	.75	–	.44	.54
Y_{1222}	.29	.54	–	.22	.47	.78	–	.37	
Y_{2222}	.35	.61	–	.25	.50	.75	–	.49	.61
Y_{1223}	.27	.51	–	.22	.47	.78	–	.33	
Y_{2223}	.36	.62	–	.25	.50	.75	–	.52	.61
Nonref. Method 3 (IIAT)									
Y_{1231}	−.05	.51	.56	.01	.10	.45	.54	.57	
Y_{2231}	−.03	.53	.52	.00	.00	.46	.54	.49	.77
Y_{1232}	−.05	.53	.58	.01	.10	.45	.54	.63	
Y_{2232}	−.03	.47	.59	.00	.00	.45	.55	.63	.77
Y_{1233}	−.05	.54	.58	.01	.10	.45	.54	.62	
Y_{2233}	−.03	.54	.59	.00	.00	.46	.54	.64	.77
Nonref. Method 4 (SR-INT)									
Y_{1241}	.40	.79	.15	.20	.45	.77	.03	.80	
Y_{2241}	.48	.75	.13	.29	.54	.69	.02	.81	.89
Y_{1242}	.42	.83	.16	.20	.45	.77	.03	.89	
Y_{2242}	.50	.77	.14	.29	.54	.69	.02	.86	.93
Y_{1243}	.42	.84	.16	.20	.45	.77	.03	.90	
Y_{2243}	.51	.79	.14	.29	.54	.69	.02	.90	.95

Notes. TCon = true trait consistency; Corr = $\sqrt{TCon}$ (correlation between measures on trait level); TMSpe = true trait method specificity; OMSpe = true occasion method specificity; Rel = reliability; CRel = composite reliability. Y_{ijkl} = observed indicator (here parcels); i = indicator; j = construct; k = method; and l = occasion of measurement. Note that the coefficient of consistency, method specificity may not add up to 1.00 in the above table due to rounding errors.

coefficient was obtained by the sum of the standardized factor loadings pertaining to the trait and to the occasion specific factor: $(.67)^2 + (.23)^2 = .50$. By comparing the results of the conscientiousness as well as intelligence reference method indicators, one notices that the reference methods (i.e., COPT and NST) both assessed stable (trait-like) rather than momentary (state-like) interindividual differences of the given constructs (see Tables 4 and 5). On average, the "true" trait consistency coefficient revealed higher values for the intelligence indicators (.90) than for the conscientiousness indicators (.78). Conversely, the average amount of occasion specificity was higher for the conscientiousness indicators (.22) than for the intelligence indicators (.10).

Each of the nonreference method indicators (i.e., C62, CIAT, SR-NEO, AMT, IIAT, and SR-INT) was affected by an indicator-specific latent trait T_{ij}, a latent trait method variable TM_{ijk}, a latent occasion method specific variable OM_{jkl}, and a measurement error variable ε_{ijkl} (see Figures 1 and 2). Again, due to the definition of the MM-LST model (see Courvoisier et al., 2008), the total observed variance of the nonreference method indicators can be decomposed into four components: the trait consistency (TCon), the trait-method specificity (TMSpe), the occasion method specificity (OMSpe), and the unreliability coefficient $[1 - \text{Rel}(Y_{ijkl})]$. The trait-method specificity coefficients (TMSpe) captures the proportion of "true" variance of a nonreference method indicator that is due to stable (consistent) interindividual differences that are *not* shared with the reference method. In other words, this coefficient reflects the amount of "true" stable interindividual differences that is free of stable interindividual differences measured by the reference method, but that is specific to the particular nonreference method. Again, the different variance coefficients were obtained by the squared values of the standardized factor loadings divided by the true variance of the particular indicator (see Tables 4 and 5). For example, for the first nonreference method indicator pertaining to the C62 measure (Y_{1121}), the standardized trait factor loading was .05, the standardized trait method factor loading was .63, and the standardized occasion specific method factor loading was .21. Hence, the "true" trait consistency coefficient

Table 6. Variances and latent correlations among the latent trait and the latent trait method factors

Construct	Conscientiousness		Intelligence		Conscientiousness						Intelligence				
Latent factors	T_{11} (COPT)	T_{21} (COPT)	T_{12} (NST)	T_{22} (NST)	TM_{112} (C62)	TM_{212} (C62)	TM_{113} (CIAT)	TM_{213} (CIAT)	TM_{114} (SR-NEO)	TM_{214} (SR-NEO)	TM_{122} (AMT)	TM_{222} (AMT)	TM_{23} (IIAT)	TM_{124} (SR-INT)	TM_{224} (SR-INT)
T_{11}	0.04														
T_{21}	.96*	0.04													
T_{12}	.51*	.40*	0.02												
T_{22}	.53*	.43*	.99*	0.02											
TM_{112}	–	–	–	–	0.03										
TM_{212}	–	–	–	–	.29*	0.02									
TM_{113}	–	–	–	–	−.06	.05	0.03								
TM_{213}	–	–	–	–	−.11	−.12	.79*	0.02							
TM_{114}	–	–	–	–	.01	−.05	.07	−.02	0.41						
TM_{214}	–	–	–	–	−.02	−.06	−.03	−.02	.73*	0.29					
TM_{122}	–	–	–	–	−.08	.10	.02	.20*	−.05	−.09	0.01				
TM_{222}	–	–	–	–	.07	.04	−.02	.09	−.07	−.16*	.77*	0.01			
TM_{23}	–	–	–	–	−.09	−.08	.61*	.96*	.02	−.05	.04	−.02	0.03		
TM_{124}	–	–	–	–	.19*	−.06	−.05	.03	.14	.10	.09	.04	.11	0.35	
TM_{224}	–	–	–	–	.12	−.06	−.05	−.01	.11	.07	.01	−.06	.00	.83*	0.40

Notes. $^*p < .05$. Variances of the latent factors are given in the diagonal.

for the first indicator was $(.05)^2 / ((.05)^2 + (.63)^2 + (.21)^2) = .01$, the trait method specificity equals $(.63)^2/((.05)^2 + (.63)^2 + (.21)^2) = .89$, and the occasion specificity coefficient equals $(.21)^2/((.05)^2 + (.63)^2 + (.21)^2) = .11$. That means that 1% of "true" variance of the nonreference method indictor (i.e., Y_{1121}) was determined by stable interindividual differences as measured by the reference method. However, the major part of the "true" variance of the indicator (i.e., 90%) was *not* determined by interindividual differences of the reference method, and was specific to stable interindividual difference measured by the nonreference method (C62 measure). Accordingly, the convergent validity between the two OPTs (COPT and C62) for the assessment of conscientiousness was low on the trait level. The correlation between both OPT measures (COPT and C62) on the trait level can be computed by taking the square root of the true trait consistency coefficient: $\sqrt{.01} = .10$ (see Table 4). With regard to the second indicator (i.e., Y_{2121}), the true trait consistency coefficient was slightly higher (6%), which corresponds to a correlation of $r = .23$ between both objective personality measures (COPT and C62). Overall, the "true" trait method specificity explains most of the true variance of the nonreference method indicators pertaining to the C62 measures (Y_{1121}–Y_{2123}), with values ranging from .81 to .90. This indicates that the two OPTs (i.e., COPT and C62) assessed different facets of conscientiousness on the trait level and, therefore, did not share much variance (i.e., low convergent validity on the trait level). However, the nonreference indicators (Y_{1121}–Y_{2123}) assessed trait-like (stable) rather than state-like (momentary) components of conscientiousness. This finding is deduced from comparing the occasion method specificity coefficient (OMSpe) with the sum of the trait consistency (TCon) and the trait method specificity (TMSpe) coefficient. With respect to the first nonreference method indicator (Y_{1121}), it follows: .01 (TCon) + .89 (TMSpe) > .10 (OMSpe). Interestingly, this relationship changed when considering the nonreference method indicators of the indirect measures (i.e., CIAT). For example, with regard to the first indicator of the indirect measures (i.e., CIAT), the trait coefficients (TCon + TMSpe) were smaller than the occasion method specificity coefficient: $.02 + .44 < .54$. This indicates that the indirect measures of conscientiousness (CIAT) assess *momentary* (occasion specific) rather than stable (consistent) interindividual differences. Nonetheless the convergent validity of the indirect measures with regard to the reference method (COPT) remained low on the trait level with trait consistencies amounting to 2% and 4%, respectively. Again, these true consistency coefficients correspond to small correlation coefficients ranging between .14 and .19 (see Tables 4 and 5). For the direct measures (e.g., SR-NEO), the trait coefficients (TCon + TMSpe) were higher than the occasion method specificity coefficients ($.00 + .96 > .04$). Again, this means that the direct measures (SR-NEO) assess stable (trait) rather than momentary components of conscientiousness. However, according to the results presented in Table 4, the trait consistency coefficients were revealed to be zero for the SR-NEO indicators, which means that self-assessed conscientiousness does not overlap with objectively assessed conscientiousness on the trait level.

Similar results were found regarding the nonreference method indicators assessing general intelligence (i.e., AMT, IIAT, SR-INT). According to the results presented in Table 5, the trait method specificity (TMSpe) coefficients as well as the occasion method specificity (OMSpe) coefficients were always large compared to the trait consistency coefficients (TCon), indicating substantial method effects on the trait as well as the occasion specific level. However, relatively high convergent correlations between both OPTs (NST and AMT) on trait level were found, ranging from .47 to .50 (see Table 5). Moreover, high convergent correlations were found between the NST and the SR-INT measures on the trait level, with values ranging between .45 and .54. This indicates that the convergent validity is higher for the intelligence measures than for the conscientiousness measures. Nevertheless, a substantial part of true variance of the nonreference method intelligence indicators could not be explained by the reference method, and this was specific to nonreference method influences. In total, the trait method specificity (TMSpe) ranged from .45 to .78. Again, the occasion method specificity (OMSpe) coefficients were low for the first OPT (NST) and the self-report measures. For the AMT, no occasion specific variance could be estimated. Therefore, it was fixed to zero. Hence, intelligence assessed by the second OPT (i.e., AMT) was not substantially affected by momentary influences. In contrast to this, the indirect measures (i.e., IIAT) assessed momentary rather than stable influences. Again, the amount of trait coefficients (TCon + TMSpe) were slightly lower than the amount of occasion method specificity (e.g., Y_{1231}: .01 + .45 < .54).

The reliability coefficients of the single indicators were comparably low except for the direct (self-report) measures (see Tables 4 and 5). However, given that it is not very common to report reliabilities for single subscales (indicators), we calculated the composite reliabilities for the entire scales according to the following formula (see also Allen, 1974):

$$\mathrm{CRel} = \frac{\sum_{i=1}^{I} \frac{\mathrm{Rel}(Y_{ijkl})}{1-\mathrm{Rel}(Y_{ijkl})}}{1 + \sum_{i=1}^{I} \frac{\mathrm{Rel}(Y_{ijkl})}{1-\mathrm{Rel}(Y_{ijkl})}}. \tag{5}$$

The composite reliabilities are also given in Tables 4 and 5. Again, the composite reliabilities are also relatively low, except for the direct (self-report) measures.

Discriminant Correlations and Generalizability of Method Effects

The correlations among the latent trait and latent trait method factors are given in Table 6. The correlations between two indicator-specific latent trait factors (e.g., T_{11} and T_{21}) belonging to the same construct j (e.g., conscientiousness), but different indicators i and i' (i.e., parcels), represent the degree of homogeneity of the indicators on the trait level. According to the results (see Table 6), the correlations between the indicator-specific trait factors for the same construct were above .95 in all cases. Hence, the indicators (i.e., parcels) were rather homogenous on the trait level. Correlations between indicator-specific latent trait factors (e.g., T_{11} and T_{12}) pertaining to the same indicator i, but to the different construct j and j' can be interpreted as the degree of discriminant validity on the trait level. High correlations between the latent trait factors indicate low discriminant validity on the trait level. According to the results in Table 6, the correlation coefficients among the latent trait factors belonging to different constructs were between 0.40 and 0.53. These relatively high positive correlations revealed low discriminant validity of the OPTs indicators on the trait level. The correlations between the latent trait method factors pertaining to the same construct j, same methods k, but different indicators i and i' (here: parcels) can be interpreted as homogeneity of method effects across indicators. High correlation coefficients indicate that the method effects generalize across indicators, whereas low correlation coefficients show that the method effects are rather indicator-specific. In our application, the correlation coefficients between these indicator-specific latent trait method factors ranged between $r_{(\mathrm{TM}_{112},\mathrm{TM}_{212})} = .29$ and $r_{(\mathrm{TM}_{123},\mathrm{TM}_{223})} = 1.00$. The first correlation coefficient shows that the latent trait method effects assessed by the C62 measure was rather indicator-specific and did not generalize across the different indicators. The latter correlation coefficient shows that the latent trait method effects assessed by the IIAT measures generalized perfectly across different indicators. Hence, a common (to all indicators) latent trait method factor (i.e., TM_{23}) was assumed instead. The correlations between the remaining latent trait method factors pertaining to different indicators were in between, ranging from $r = .73$ to $r = .83$. This indicates that the latent trait method effects assessed by the remaining methods were rather homogenous across different indicators. However, given that these method effects did not generalize perfectly across different indicators, it was necessary to specify indicator-specific latent trait method factors instead of common latent trait method factors.

The correlations between latent trait method factors pertaining to the same construct j but different methods k and k' (where $k, k' \neq 1$) can be interpreted as degree of generalizability of method effects. For example, high positive correlations would indicate that method specific deviations from the values predicted by the latent trait factor as measured by the reference method are associated with one another. According to Table 6, all of these correlations were insignificant. Hence, the latent trait method effects assessed by one method (e.g., direct measures) were not related with latent trait method effects assessed by another method (e.g., indirect measures) measuring the same construct.

The correlations between latent trait method factors pertaining to different constructs can be interpreted in a similar way. These correlations indicate whether or not latent trait method factors generalize across different constructs and/or across different methods on the trait level. The highest correlations in our application were found among the latent trait method factors belonging to the indirect measure (CIAT, IIAT) of conscientiousness and intelligence (see Table 6). These correlations indicate that the latent trait method effects of the indirect measures generalize across both constructs. In other words, indirect measures (CIAT, IIAT) share something in common in their assessment of conscientiousness and intelligence that is not shared with the reference methods (COPT, NST). The correlations between the remaining latent trait method factors pertaining to different constructs j and j' and different methods k and k' (where k, $k' \neq 1$) were rather low and ranged between $r = -.16$ and $r = .20$.

Finally, the discriminant validity as well as the generalizability on the momentary (occasion specific) level was not directly calculated, given that the amount of occasion specific variability was too low to allow the estimation of the correlations.

Criterion Validity of OPTs

In Table 7, the results of the latent regressions of the behavioral outcome of conscientiousness on the latent trait factors are given. As a criterion for conscientiousness, we measured the punctuality (in minutes) of the arrival to the test session 2 and session 3. The general punctuality of the test takers was assessed by specifying another latent trait factor for the two punctuality measures at session 2 and session 3. Since the indicator-specific latent trait as well as latent trait method factors were highly correlated (see Table 6), we conducted separate latent regression analyses for each indicator (parcel). The results show that the latent trait factors (T_{11} and T_{21}), as measured by the reference method, are negatively related with the criterion (i.e., punctuality). This finding indicates that a tendency to arrive earlier to the test sessions than the fixed time was found in individuals with higher conscientiousness scores on the reference trait factors.[4] Specifically, persons who differed by one unit with respect to their conscientiousness level measured by T_{11} and T_{21}, in general tended to arrive -5.96 and -5.97 min earlier to the test sessions. Moreover, the latent trait method factor (TM_{114} and TM_{214}) was also negatively related to the criterion. Again, the latent trait method factors (TM_{114} and TM_{214}) represent the amount of stable (consistent) self-reported conscientiousness that is not shared with the stable trait as measured by the reference method, but that is specific for the nonreference method (i.e., self-report measures). Hence, the stable self-report of an individual's conscientiousness that cannot be predicted by the reference method was also predictive for behavioral outcomes (i.e., punctuality). However, the other latent trait method factors were not significantly associated with the behavioral outcome (see Table 7).

Discussion

In this study, we investigated the reliability as well as the construct validity of objective personality tests (OPTs) assessing conscientiousness and intelligence across time. Specifically, we scrutinized the convergent, discriminant, and criterion validity of OPTs across time by using a longitudinal MTMM structural equation model, the so-called MM-LST model (Courvoisier et al., 2008). Furthermore, we gained information on the reliability of different OPTs and added data on the criterion validity of the OPTs employed to assess conscientiousness. The results of the analysis revealed new insights regarding the validity of OPTs.

First, convergent correlations between OPTs and other measures assessing conscientiousness were overall low. More specifically, the latent correlations among the different measures (i.e., convergent validity on the trait level) ranged from 0 to .24. Zero correlations were found between direct and objective personality measures, indicating no convergent validity between these measures on the trait level. Somewhat higher convergent correlations were found on the trait level for the assessment of general intelligence among the different measures, ranging from 0 to .54. Interestingly, the convergent correlations between self-reports (direct) and objective intelligence measures were relatively high, ranging between $r = .45$ and $r = .54$. This means that the convergent validity on the trait level between direct measures (self-reports) and OPTs was higher for the assessment of general intelligence than for the assessment of conscientiousness. The convergent correlations among the OPTs were low for conscientiousness ($r = .10$–$.24$) and medium for intelligence ($r = .47$–$.50$). The convergent correlations between the OPTs and the indirect measures were low for both constructs, ranging between 0 and .20.

Second, the findings of our study suggest that the OPTs we employed assessed stable (trait-like) rather than momentary (state-like) components of conscientiousness and intelligence. The MM-LST model by Courvoisier et al. (2008) enables researchers to examine the convergent validity on the momentary (occasion specific) as well as on the stable (trait) level. According to our findings, the measures used in the study assessed only a small amount of occasion specific influences, with the exception of the indirect measures (i.e., CIAT and IIAT). In our study, the indirect measures showed almost equally sized coefficients of occasion method specificity and of trait method specificity, indicating that the indirect measures are influenced by stable trait and occasion specific influences (i.e., occasion specific deviations from the time-invariant trait) to the same degree. For the remaining measures (OPTs, self-report scales), the trait consistency coefficients (TCon for the reference method; TCon + TMSpe for the remaining methods) exceeded the occasion specificity coefficients, suggesting that these measures assessed traits rather than state-like constructs. Due to the small amount of occasion specific effects, the

[4] Following a recommendation of a reviewer, we rerun the analysis using the self-reports as reference method. In both analyses, the OPT (COPT) and the self-report measure (SR-NEO) predicted punctuality. Thus, changing the reference method had no effect on the results of the predictive validity of the COPT or the SR-NEO measures.

Table 7. Latent regression analysis for the 1st item parcel (punctuality on trait and trait method factors)

	Model 1 (1st Item Parcel)			Model 2 (2nd Item Parcel)		
Factors	Model (*B*)	Model (β)	Model (*SE*)	Model (*B*)	Model (β)	Model (*SE*)
Constant	−1.52			−1.53		
Trait factor (COPT)	−5.96***	−.26	2.31	−5.97***	−.27	2.30
1. Trait Method Factor (C67)	3.91	.16	2.55	−4.56	−.15	3.15
2. Trait Method Factor (CIAT)	−.07	.00	4.07	−2.54	−.08	4.65
3. Trait Method Factor (SR-NEO)	−1.83***	−.27	.61	−3.16***	−.40	.72
R^2	.17			.25		

Notes. ****p* < .001. Punctuality in terms of arriving early, on time, or late to the scheduled appointed measured in minutes.

convergent validity on the momentary (occasion specific) level was zero for all measures.

Third, according to our findings, the composite (total scale) reliabilities of the measures were low to excellent. The direct measures (i.e., self-report questionnaires) revealed the highest (i.e., excellent) reliability coefficients. Two of our four OPTs (i.e., C62, AMT) had rather low reliability coefficients.

Fourth, our findings suggest that OPTs and the remaining (direct and indirect) measures assessed different facets of conscientiousness and general intelligence. This conclusion can be derived from the substantial method effects of the nonreference methods that were observed for all methods on the trait level and for the indirect measures on the occasion specific level as well. Trait method specificity coefficients ranged from .37 to .96, occasion method specificity coefficients from .02 to .59. These results indicate that the different methods assess unique components of conscientiousness and intelligence that are not shared with the reference method. Hence, the methods we compared cannot be interchanged with one another. Instead, they complement each other in that they capture facets of the measured constructs that cannot be assessed by other methods.

Fifth, we examined the discriminant validity among the measures as well as the generalizability of method effects. The discriminant validity was rather low on the trait level as measured by the reference method, revealing substantial correlations between the latent traits across constructs ranging from .40 to .53. Within each construct, the method effects did not generalize across different methods indicating that the method effects are unique. For example, the method specific components of self-reports did not correlate with the method specific components of indirect measures. However, method effects did generalize across different constructs in some cases. This generalizability of method effects was most pronounced for the indirect measures on the trait level. More specifically, the correlation between the two latent trait method factors of the two conscientiousness IATs had correlations of .61 and .93 with the single latent trait method factor of the two intelligence IATs (cf. Table 6). This clearly indicates that the IAT has limited validity because it not only measure the specific construct that it was designed to measure but irrelevant common factors such as recoding strategies that generalize across IATs for different constructs as well.

Sixth, we scrutinized the criterion validity of the conscientiousness measures by relating the latent trait as measured by the reference method as well as the trait method factors to punctuality on the second and third test sessions as a relevant behavioral outcome of conscientiousness. According to the results of these latent regression analyses, the latent trait as measured by the reference OPT (COPT) as well as the latent trait method self-report factor (SR-NEO) predicted punctuality. These results support the claim that OPTs and self-reports assess different facets of conscientiousness which are both predictive for real life criteria.

In sum, from the perspective of the "classical" MTMM rationale that requires correlations between different methods aiming to assess the same construct to be higher as compared to correlations between the same methods for assessing different constructs, the results from our analyses are disappointing. This is especially true for the two conscientiousness OPTs that we employed. These OPTs showed little convergence with each other and with other conscientiousness measures. Unlike the conscientiousness OPTs, the intelligence tests converged substantially among each other and with self-reports. Limited construct validity according to the MTMM rationale was also observed for the IATs that had high method specificity both on the trait level and on the occasion specific level. Moreover, the IATs lacked discriminant validity on the trait level as their method effects generalized across constructs.

The MTMM rationale may not, however, always be the only and sometimes even not the best rationale for estimating the construct validity of measures. This is true because the MTMM rationale starts out from the presumption that different methods have to measure identical constructs. This presumption may not always be appropriate as the "method-trait problem" suggests. Rather, different methods may sometimes capture unique components of constructs that cannot be assessed with other methods. This issue was discussed extensively by Cattell and colleagues (Hundleby et al., 1965) and once again more recently with regard to the low convergence between direct and indirect measures (Gschwendner et al., 2008). Several authors have claimed that direct self-report methods and indirect methods such as the IAT assess different constructs, that is, explicit dispositions and implicit dispositions (Wilson, Lindsey, & Schooler, 2000). Explicit and implicit dispositions are related but not identical. This claim is supported by findings showing that direct and indirect measures have incremental validity in predicting behavioral outcomes (Friese et al., 2008; Greenwald et al., 2009). Our results regarding the predictability of punctuality as a behavioral outcome of conscientiousness are consistent with this view. Punctuality was predicted by the latent conscientiousness trait as measured by the reference OPT (COPT) and by the latent trait method self-report factor (SR-NEO).

The view that different methods measure unique components of constructs has important implications. Most importantly perhaps, it implies that the existence of method effects does not necessarily mean that these effects are exclusively due to systematic measurement error that reduces the validity of the method. Rather, methods effects may reflect, at least to some extent, substantively relevant variance, that is, a component of the construct to be measured (Schmitt, 2006).

What alternative or additional strategies for estimating construct validity do we have in such cases? We propose that the application of a substantive theory is the most convincing strategy. Construct validity is given to the extent that a measurement method generates data that are in agreement with a theory in which the construct to be measured is included. Schmitt, Hofmann, Gschwendner, Gerstenberg, and Zinkernagel (in press) have proposed a theoretical model that might be useful for the simultaneous analysis of the construct validity of the types of methods we have looked at in the present study (OPT, IAT, self-report). In line with dual process theories (Strack & Deutsch, 2004), the Schmitt et al. model assumes (a) that explicit dispositions can be measured directly with self-report scales, (b) that implicit dispositions can only be measured indirectly with procedures like the IAT, (c) that explicit dispositions affect behavior via plans and intentions, and (d) that implicit dispositions affect behavior via the automatic activation of behavioral scripts and schemata. Going beyond classical dual process theories, the model assumes that these effects are moderated by personality and situation factors. It is assumed, for example, that the effects of explicit and implicit dispositions on behavior depend on self-control resources that are available when behavior is being executed. If self-control is high (low), explicit dispositions will be more (less) influential as compared to implicit dispositions. This moderator effect can be traced both to interindividual differences in self-control resources (Hofmann, Gschwendner, Wiers, Friese, & Schmitt, 2008) and to situational differences (Hofmann, Gschwendner, Castelli, & Schmitt, 2008) or both. The Schmitt et al. (in press) model has received empirical support in several studies (for overviews see Friese et al., 2008; Schmitt et al., in press) and has been applied to OPTs in two studies (Dislich et al., 2010). In these studies, the effect of implicit and explicit risk propensity on risk behavior as measured by risk OPTs depended, as predicted, on the impulsiveness of the person and the impulsiveness of the OPT. This type of theory-driven research is both necessary and promising in addition to the conventional MTMM research that we have reported in the present paper.

Conclusion

According to the results of this study, the OPT measure trait-like rather than state-like attributes. The OPT are relatively reliable, however, in comparison to the self-report measures an increase of the internal consistencies of the OPTs would be advantageous. The OPTs showed higher convergent validity coefficients with other OPTs measures than with implicit measures, and in the case of conscientiousness also with self-report measures. These results seem reasonable, given that these different types of methods (self-reports, implicit measures, and OPT) seem to assess fundamentally different facets of the same attribute (here: conscientiousness and general intelligence). The discriminant validity of the OPTs revealed sufficient, given that OPTs can differentiate between both constructs. Finally, the OPTs were significantly related to behavioral outcome variables (here: punctuality).

Advantages and Limitations of the Study

Finally, we address the advantages as well as limitations of the present study. One main advantage of this study is that the convergent and discriminant validity of OPTs and other (direct and indirect) measures were investigated across time by using a longitudinal MTMM-MO approach. In particular, the MM-LST model allows researchers to scrutinize the "true" (measurement error-free) convergent and discriminant validity of the given measures at each occasion of measurement as well as across occasions of measurement. Moreover, the MM-LST model allows disentangling the consistent (stable) interindividual differences from momentary (occasion specific) interindividual differences and thereby allows for the investigation of convergent and discriminant validity on the trait as well as occasion specific level. Given that the MM-LST model makes use of a "regression type" [i.e., CTC($M-1$)] approach, it is possible to contrast different methods against a reference (golden standard) method, even if these methods use different scales (see also Geiser et al., 2012). This proved to be particularly beneficial in our study, where conscientiousness and intelligence were measured using a mixture of methods (OPTs, indirect measures, and direct measures) using different scales.

However, the present study is also limited in several ways. First, it was not feasible to conduct the analysis on the item level due to the large amount of items (sometimes more than 20) per CMOU. Consequently, it was necessary to reduce the complexity of the model and use parcels (test halves) instead of single items. Second, in this study we did not incorporate or manipulate manifest or latent moderator variables with regard to the convergent validity on the trait and state level. However, this might be useful when testing the theoretical model proposed by Schmitt et al. (in press). Finally, future studies should attempt to include OPTs assessing additional constructs and a larger variety of covariates that may explain the trait and/or the method effects on the trait as well as occasion specific level. Finally, future research should broaden this approach by including additional personality constructs and more OPTs in order to test whether the results presented here can be generalized both across different constructs and across different OPTs.

Acknowledgments

We would like to express our great appreciation to Markus Meindl, Juliane Caspers, Eva Ebenführer, Tobias Stadler, Elisabeth Becker, Benjamin Augustin, Nicole Mathwig, Franziska Vogel, and Lisa Grünberger for their support and help in the project. The project was founded by the German Research Foundation (DFG, Project Number OR 238\3-1).

References

Allen, M. P. (1974). Construction of composite measures by the canonical factor regression method. In H. L. Costner (Ed.), *Sociological methodology 1973–1974* (pp. 51–78). San Francisco, CA: Jossey-Bass.

Amthauer, R., Brocke, B., Liepmann, D., & Beauducel, A. (2001). *Intelligenz-Struktur-Test 2000 R* [Test of intelligence structure 2000 R]. Göttingen, Germany: Hogrefe.

Back, M. D., Krause, S., Hirschmuller, S., Stopfer, J. M., Egloff, B., & Schmukle, S. C. (2009). Unraveling the three faces of self-esteem: A new information-processing sociometer perspective. *Journal of Research in Personality, 43*, 933–937. doi: 10.1016/j.jrp.2009.04.002

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*, 238–246.

Bentler, P. M. (1995). *EQS structural equations program manual*. Encino, CA: Multivariate Software.

Bentler, P., & Chou, C.-P. (1987). Practical issues in structural modeling. *Sociological Methods & Research, 17*, 78–117.

Bosson, J. K., Brown, R. P., Zeigler-Hill, V., & Swann, W. B. Jr. (2003). Self-enhancement tendencies among people with high self-esteem: The moderating role of implicit self-esteem. *Self and Identity, 2*, 169–187. doi: 10.1080/15298860309029

Brand, M., Fujiwara, E., Borsutzky, S., Kalbe, E., Kessler, J., & Markowitsch, H. J. (2005). Decision-making deficits of Korsakoff patients in a new gambling task with explicit rules: Associations with executive functions. *Neuropsychology, 19*, 267–277.

Brand, M., Heinze, K., Labudda, K., & Markowitsch, H. J. (2008). The role of strategies in deciding advantageously in ambiguous and risky situations. *Cognitive Processing, 9*, 159–173.

Brickenkamp, R. (2002). *Test d2 Aufmerksamkeits-Belastungs-Test. Manual. 9., überarbeitete und neu normierte Auflage* [The d2 Test of Attention. Manual] (9th ed.). Göttingen, Germany: Hogrefe.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81–105.

Carroll, J. B. (1993). *Human cognitive abilities*. New York, NY: Cambridge University Press.

Cattell, J. M. (1890). Mental tests and measurements. *Mind, 15*, 373–381.

Cattell, R. B. (1941). An objective test of character-temperament: I. *Journal of General Psychology, 25*, 59–73.

Cattell, R. B. (1944). An objective test of character-temperament: II. *Journal of Social Psychology, 19*, 99–114.

Cattell, R. B. (1946). *Description and measurement of personality*. New York, NY: World Book.

Cattell, R. B. (1957). *Personality and motivation structure and measurement*. New York, NY: World Book.

Cattell, R. B. (1968). What is "objective" in "objective personality tests"? In W. L. Barnette (Ed.), *Readings in psychological tests and measurements* (pp. 163–168). Homewood, IL: Dorsey Press.

Cattell, R. B., & Kline, P. (1977). *The scientific analysis of personality and motivation*. London, UK: Academic Press.

Cattell, R. B., & Warburton, F. W. (1967). *Objective personality and motivation tests: A theoretical introduction and practical compendium*. Chicago, IL: University of Illinois Press.

Costa, P. T., & McCrae, R. R. (1992). *Neo PI-R professional manual*. Odessa, FL: Psychological Assessment Resources.

Courvoisier, D. S. (2006). *Unfolding the constituents of psychological scores: Development and application of mixture and multitrait-multimethod LST models* (Doctoral dissertation). University of Geneva, Switzerland. Retrieved from http://doc.rero.ch/record/9221/files/these.pdf

Courvoisier, D. S., Nussbeck, F. W., Eid, M., & Cole, D. A. (2008). Analyzing the convergent and discriminant validity of states and traits: Development and applications of multimethod latent state-trait models. *Psychological Assessment, 20*, 270–280. doi: 10.1037/a0012812

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281–302.

Dislich, F. X. R., Imhoff, R., Banse, R., Altstötter-Gleich, C., Zinkernagel, A., & Schmitt, M. (2012). Discrepancies between implicit and explicit self-concepts of intelligence predict performance on tests of intelligence. *European Journal of Personality, 26*, 212–220.

Dislich, F. X. R., Zinkernagel, A., Ortner, T. M., & Schmitt, M. (2010). Convergence of direct, indirect, and objective risk taking measures in the domain of gambling: The moderating role of impulsiveness and self-control. *Journal of Psychology, 218*, 20–27.

Egloff, B., Schwerdtfeger, A., & Schmukle, S. C. (2005). Temporal stability of the Implicit Association Test-Anxiety. *Journal of Personality Assessment, 84*, 82–88.

Eid, M. (2000). A multitrait-multimethod model with minimal assumptions. *Psychometrika, 65*, 241–261.

Eid, M., Lischetzke, T., Nussbeck, F. W., & Trierweiler, L. I. (2003). Separating trait effects from trait-specific method effects in multitrait-multimethod models: A multiple-indicator CT-C(M-1) model. *Psychological Methods, 8*, 38–60.

Eid, M., Nussbeck, F. W., Geiser, C., Cole, D. A., Gollwitzer, M., & Lischetzke, T. (2008). Structural equation modeling of multitrait-multimethod data: Different models for different types of methods. *Psychological Methods, 13*, 230–253.

Elliot, S., Lawty-Jones, M., & Jackson, C. (1996). Effects of dissimulation on self-report and objective measures of personality. *Personality and Individual Differences, 21*, 335–343.

Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and uses. *Annual Review of Psychology, 54*, 297–327.

Finney, S. J., & DiStefano, C. (2006). Non-normal and categorical data in structural equation modeling. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 269–314). Greenwich, CT: Information Age Publishing.

Friese, M., Hofmann, W., & Schmitt, M. (2008). When and why do implicit measures predict behaviour? Empirical evidence for the moderating role of opportunity, motivation, and process reliance. *European Review of Social Psychology, 19*, 285–338. doi: 10.1080/10463280802556958

Geiser, C., Eid, M., West, S. G., Lischetzke, T., & Nussbeck, F. W. (2012). A comparison of method effects in two confirmatory factor models for structurally different methods. *Structural Equation Modeling, 19*, 409–436. doi: 10.1080/10705511.2012.687658

Geiser, C., Keller, B., Lockhart, G., Eid, M., Cole, D., & Koch, T. (2014). Distinguishing state variability from trait change in longitudinal data: The role of measurement (non)invariance in latent state-trait analyses. *Behaviour Research Methods*. doi: 10.3758/s13428-014-0457-z

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.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test I: An improved scoring algorithm. *Journal of Personality and Social Psychology, 85*, 197–216.

Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology, 97*, 17–41. doi: 10.1037/a0015575

Gschwendner, T., Hofmann, W., & Schmitt, M. (2008). Differential stability: The effects of acute and chronic construct accessibility on the temporal stability of the Implicit Association Test. *Journal of Individual Differences, 29*, 70–79. doi: 10.1027/1614-0001.29.2.70

Häcker, H., Schmidt, L. R., Schwenkmezger, P., & Utz, H. E. (1975). *Objektive Testbatterie, OA-TB 75* [Objective Test Battery, OA-TB 75]. Weinheim, Germany: Beltz.

Häcker, H., Schwenkmezger, P., & Utz, H. (1979). Über die Verfälschbarkeit von Persönlichkeitsfragebogen und Objektiven Persönlichkeitstests unter SD-Instruktion und in einer Auslesesituation [About fakeability of personality questionnaires under SD-instruction and in a situation of selection]. *Diagnostica, 25*, 7–23.

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the implicit association test and explicit self-report measures. *Personality and Social Psychology Bulletin, 31*, 1369–1385. doi: 10.1177/0146167205275613

Hofmann, W., Gschwendner, T., Castelli, L., & Schmitt, M. (2008). Implicit and explicit attitudes and interracial interaction: The moderating role of situationally available control resources. *Group Processes & Intergroup Relations, 11*, 69–87.

Hofmann, W., Gschwendner, T., Wiers, R., Friese, M., & Schmitt, M. (2008). Working memory capacity and self-regulatory behavior: Towards an individual differences perspective on behavior determination by automatic versus controlled processes. *Journal of Personality and Social Psychology, 95*, 962–977.

Hornke, L. F., Küppers, A., & Etzel, S. (2000). Design and evaluation of an adaptive matrices test. *Diagnostica, 46*, 182–188.

Hu, L.-T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3*, 424–453. doi: 10.1037/1082-989x.3.4.424

Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1–55. doi: 10.1080/10705519909540118

Hundleby, J. D., Pawlik, K., & Cattell, R. B. (1965). *Personality factors in objective test devices*. San Diego, CA: Knapp.

Karpinski, A., & Steinman, R. B. (2006). The single category implicit association test as a measure of implicit social cognition. *Journal of Personality and Social Psychology, 91*, 16–32.

Kubinger, K. D., & Ebenhöh, J. (1996). *Arbeitshaltungen – Kurze Testbatterie: Anspruchsniveau, Frustrationstoleranz, Leistungsmotivation, Impulsivität/Reflexivität* [Working Style – A short test-battery: Level of aspiration, achievement motivation, frustration tolerance, impulsiveness/reflexiveness] [Software and Manual]. Frankfurt/M.: Swets Test Services.

Lejuez, C. W., Aklin, W. M., Jones, H. A., Richards, J. B., Strong, D. R., Kahler, C. W., & Read, J. P. (2003). The balloon analogue risk task (BART) differentiates smokers and nonsmokers. *Experimental and Clinical Psychopharmacology, 11*, 26–33.

Lejuez, C. W., Read, J. P., Kahler, C. W., Richards, J. B., Ramsey, S. E., Stuart, G. L., ... Brown, R. A. (2002). Evaluation of a behavioral measure of risk taking: The Balloon Analogue Risk Task (BART). *Journal of Experimental Psychology-Applied, 8*, 75–84. doi: 10.1037/1076-898x.8.2.75

Lischetzke, T., Eid, M., & Nussbeck, F. W. (2002, September). *Unterschiedliche Definitionen von Methodeneffekten in MTMM Modellen und ihre Implikationen für die Analyse der Validität* [Different definitions of method effects in MTMM models and their implications for validity analysis]. Paper presented at the 43rd meeting of the German Psychological Association, Berlin, Germany.

Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling, 9*, 151–173. doi: 10.1207/s15328007sem0902_1

Matthies, S., Philipsen, A., & Svaldi, J. (2012). Risky decision making in adults with ADHD. *Journal of Behavior Therapy and Experimental Psychiatry, 43*, 938–946.

Muthén, L. K., & Muthén, B. O. (1998–2012). *Mplus user's guide* (3rd ed.). Los Angeles, CA: Muthén & Muthén.

Nussbeck, F. W., Eid, M., & Lischetzke, T. (2006). Analysing multitrait-multimethod data with structural equation models for ordinal variables applying the WLSMV estimator: What sample size is needed for valid results? *British Journal of Mathematical & Statistical Psychology, 59*, 195–213. doi: 10.1348/00071100sx67490

Ortner, T. M., Horn, R., Kersting, M., Krumm, S., Kubinger, K. D., Proyer, R. T., ... Westhoff, K. (2007). Standortbestimmung und Zukunft Objektiver Persönlichkeitstests [Current situation and future of objective personality tests]. *Report Psychologie, 32*, 64–75.

Ortner, T. M., Kubinger, K. D., Schrott, A., Radinger, R., & Litzenberger, M. (2006). *Belastbarkeits-Assessment: Computerisierte Objektive Persönlichkeits-Testbatterie – Deutsch (BAcO-D)* [Stress resistance assessment: Computerized Objective Test-battery – German version] [Test: Software und Manual]. Frankfurt/M.: Harcourt Assessment.

Ortner, T. M., & Proyer, R. T. (2014). *Objective Personality Tests.* Manuscript submitted for publication.

Ostendorf, F. (1994). Zur Taxonomie deutscher Dispositionsbegriffe [A taxonomy of German dispositional concepts]. In W. Hager & M. Hasselhorn (Eds.), *Handbuch deutschsprachiger Wortnormen* [Handbook of German word norms] (pp. 382–441). Göttingen, Germany: Hogrefe.

Ostendorf, F., & Angleitner, A. (2003). *NEO-Persönlichkeitsinventar nach Costa und McCrae, Revidierte Fassung (NEO-PI-R)* [NEO-Personality Inventory revisited by Costa and McCrae]. Göttingen, Germany: Hogrefe.

Pohl, S., & Steyer, R. (2010). Modeling common traits and method effects in multitrait-multimethod analysis. *Multivariate Behavioral Research, 45*, 45–72.
Pohl, S., Steyer, R., & Kraus, K. (2008). Modelling method effects as individual causal effects. *Journal of the Royal Statistical Society Series A, 171*, 41–63.
Proyer, R. T. (2007). Convergence of conventional and behavior-based measures: Towards a multimethod approach in the assessment of vocational interests. *Psychology Science Quarterly, 49*, 168–183.
Proyer, R. T., & Häusler, J. (2007). Assessing behavior in standardized settings: The role of objective personality tests. *International Journal of Clinical and Health Psychology, 7*, 537–546.
R Development Core Team. (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org
Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen, Denmark: Nielsen & Lydiche.
Rothermund, K., & Wentura, D. (2004). Underlying processes in the Implicit Association Test: Dissociating salience from associations. *Journal of Experimental Psychology: General, 133*, 139–165.
Rubio, V. J., Hernández, J. M., Zaldivar, F., Marquez, O., & Santacreu, J. (2010). Can we predict risk-taking behavior? Two behavioral tests for predicting guessing tendencies in a multiple-choice test. *European Journal of Psychological Assessment, 26*, 87–94.
Santacreu, J., Rubio, V. J., & Hernández, J. M. (2006). The objective assessment of personality: Cattells's T-data revisited and more. *Psychology Science, 48*, 53–68.
Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research-Online, 8*, 23–74. doi: citeulike-article-id:7222182
Schmidt, L. R. (1975). *Objektive Persönlichkeitsmessung in diagnostischer und klinischer Psychologie* [Objective personality measurement in psychological assessment and clinical psychology]. Weinheim, Germany: Beltz.
Schmidt-Atzert, L. (2004). *Objektiver Leistungsmotivations Test (OLMT)* [Objective Achievement Motivation Test] [Software and Manual]. Mödling, Austria: Dr. G. Schuhfried.
Schmidt-Atzert, L. (2007). *Objektiver Leistungsmotivations Test (OLMT)* [Objective Achievement Motivation Test] [Software and Manual]. Mödling, Austria: Dr. G. Schuhfried.
Schmitt, M. (2006). Conceptual, theoretical, and historical foundations of multimethod assessment. In M. Eid & E. Diener (Eds.), *Handbook of multimethod measurement in psychology* (pp. 9–25). New York, NY: American Psychological Association.
Schmitt, M., Hofmann, W., Gschwendner, T., Gerstenberg, F. X., & Zinkernagel, A. (in press). A model of moderated convergence between direct, indirect, and behavioral measures of personality traits. In A. J. R. Van de Vijver & T. M. Ortner (Eds.), *Behavior based assessment: Going beyond self report in the personality, affective, motivation, and social domains*. Göttingen, Germany: Hogrefe.
Schmukle, S. C., Back, M. D., & Egloff, B. (2008). Validity of the five-factor model for the implicit self-concept of personality. *European Journal of Psychological Assessment, 24*, 263–272. doi: 10.1027/1015-5759.24.4.263
Schmukle, S. C., & Egloff, B. (2005). A latent state-trait analysis of implicit and explicit personality measures. *European Journal of Psychological Assessment, 21*, 100–107. doi: 10.1027/1015-5759.21.2.100
Skinner, N. S. F., & Howarth, E. (1973). Cross-media independence of questionnaire and objective test personality factors. *Multivariate Behavioral Research, 8*, 23–40.
Steffens, M. C., & König, S. S. (2006). Predicting spontaneous big five behavior with implicit association tests. *European Journal of Psychological Assessment, 22*, 13–20. doi: 10.1027/1015-5759.22.1.13
Steiger, J. H. (1990). Structural model evaluation and modification - An interval estimation approach. *Multivariate Behavioral Research, 25*, 173–180. doi: 10.1207/s15327906mbr2502_4
Steyer, R., Schmitt, M., & Eid, M. (1999). Latent state-trait theory and research in personality and individual differences. *European Journal of Personality, 13*, 389–408.
Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review, 8*, 220–247.
Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review, 107*, 101–126.
Ziegler, M., Schmidt-Atzert, L., Bühner, M., & Krumm, S. (2007). Faking susceptibility of different measurement methods: Questionnaire, semi-projective, and objective. *Psychology Science, 49*, 291–307.
Ziegler, M., Schmukle, S. C., Egloff, B., & Bühner, M. (2010). Investigating Measures of Achievement Motivation(s). *Journal of Individual Differences, 31*, 15–21. doi: 10.1027/1614-0001/a000002

Date of acceptance: February 15, 2014
Published online: July 16, 2014

Tobias Koch

Department of Education and Psychology
Freie Universität Berlin
Habelschwerdter Allee 45
14195 Berlin
Germany
Tel. +49 30 83855719
E-mail tkoch@fedat.fu-berlin.de

Appendix

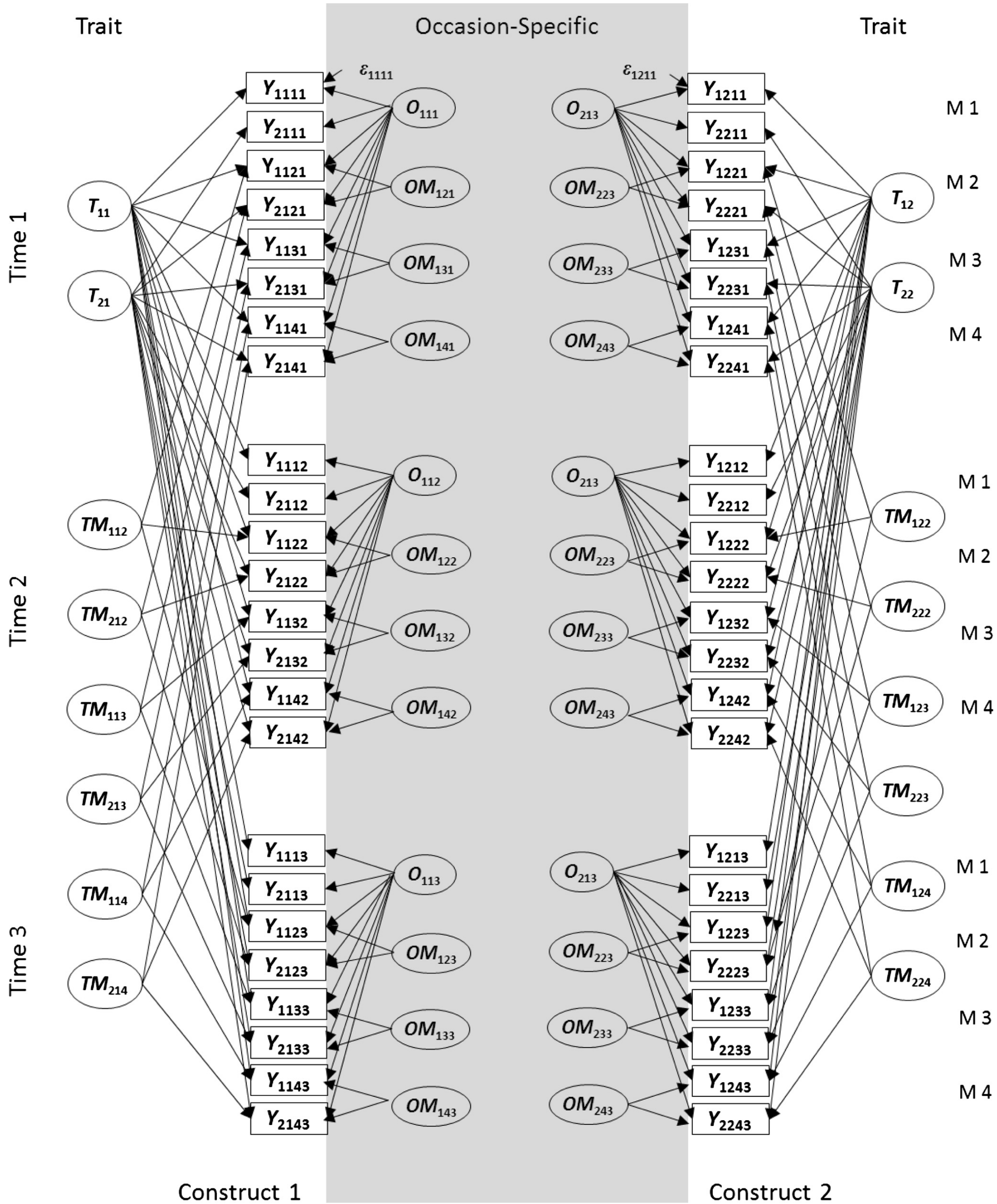

Figure A-1. Multitrait-multimethod latent state-trait (MM-LST) model (Courvoisier et al., 2008) for three indicators, two constructs, four methods, and three occasions of measurement. T_{ij} = reference trait factor (indicator-specific), TM_{ijk} = latent trait method factors (indicator-specific), O_{j1l} = reference occasion specific factors (indicator-unspecific), OM_{jkl} = occasion specific method factors (indicator-unspecific), ε_{ijkl} = measurement error variables, i = indicator, j = construct, k = method, l = occasion of measurement. For the sake of clarity, the factor loading parameters and latent correlations have been omitted in the figure.

