



Journal of Personality Assessment

ISSN: 0022-3891 (Print) 1532-7752 (Online) Journal homepage: <http://www.tandfonline.com/loi/hjpa20>

Improving the Assessment of Implicit Motives Using IRT: Cultural Differences and Differential Item Functioning

J. Malte Runge, Jonas W. B. Lang, Athanasios Chasiotis & Jan Hofer

To cite this article: J. Malte Runge, Jonas W. B. Lang, Athanasios Chasiotis & Jan Hofer (2018): Improving the Assessment of Implicit Motives Using IRT: Cultural Differences and Differential Item Functioning, Journal of Personality Assessment, DOI: [10.1080/00223891.2017.1418748](https://doi.org/10.1080/00223891.2017.1418748)

To link to this article: <https://doi.org/10.1080/00223891.2017.1418748>



Published online: 01 Feb 2018.



Submit your article to this journal [↗](#)



Article views: 29



View related articles [↗](#)



View Crossmark data [↗](#)



Improving the Assessment of Implicit Motives Using IRT: Cultural Differences and Differential Item Functioning

J. Malte Runge,¹ Jonas W. B. Lang,¹ Athanasios Chasiotis,² and Jan Hofer³

¹Department of Personnel Management, Work and Organizational Psychology, Ghent University, Ghent, Belgium; ²Department of Developmental Psychology, Tilburg University, Tilburg, The Netherlands; ³Department of Developmental Psychology, Trier University, Trier, Germany

ABSTRACT

Researchers have long been interested in studying differences in implicit motive between different groups. Implicit motives are typically measured by scoring text that respondents have written in response to picture cues. Recently, research on the measurement of implicit motives has made progress through the application of a dynamic Thurstonian item-response theory model (DTM; Lang, 2014) that captures 2 basic motivational processes in motivational research: motive competition and dynamic reduction of motive strength after a motive has been acted out. In this article, the authors use the DTM to investigate differential item functioning (DIF) in implicit motive measures. The article first discusses DIF in the context of the DTM. The authors then conduct a DIF analysis of data from a study that used a picture set of the Operant Motive Test (OMT; Kuhl & Scheffer, 2002) with participants from Cameroon, Germany, and Costa Rica. Results showed no evidence of DIF in 9 pictures and some evidence for DIF in 3 pictures. The authors show a partial invariance model can be specified and use this partial invariance model to study latent mean differences between Cameroon, Germany, and Costa Rica. The discussion focuses on the use of IRT DIF methods in future research on implicit motives.

ARTICLE HISTORY

Received 19 May 2017
Revised 17 November 2017

Motivational researchers have long been interested in cultural differences (and similarities) in the development of implicit motives (Chasiotis, Bender, & Hofer, 2014) and cultural differences in the link between implicit motives and behavioral outcomes (e.g., Hofer & Bond, 2008; Hofer, Kärtner, Chasiotis, Busch, & Kiessling, 2007). One core challenge in cultural group comparisons using implicit motive measures is that respondents from different cultural groups might differ in the way they interpret picture cues. Differences between cultures in implicit motive measures would then not result from real differences in implicit motives, but from cultural bias in picture perception. A methodological tool designed to distinguish between real existing group differences (impact) and test bias (Thissen, Steinberg, & Gerrard, 1986) is differential item functioning (DIF). The core goal of DIF is to identify response options that function differently across groups.

DIF is typically studied using item-response theory (IRT). However, IRT-DIF analyses require a viable IRT measurement model for the psychological individual differences construct of interest. The response process in implicit motive measures is complex because the motives do not act independently (Atkinson & Birch, 1970). Recently, researchers developed an IRT measurement model—the dynamic Thurstonian IRT model (DTM; Lang, 2014)—that accounts for the dependency of the motives and applied variants of this model to implicit motive measures (Lang, 2014; Lang, Zettler, Ewen, & Hülshager, 2012; Runge et al., 2016). This measurement model builds on recent advancements in the

modeling of forced-choice data using Thurstonian IRT (Brown & Maydeu-Olivares, 2013) and additionally accounts for a dynamic reduction in motive strength after motive enactment (Tuerlinckx, De Boeck, & Lens, 2002; Verhelst & Glas, 1993).

This article seeks to contribute to the literature in two ways. First, we contribute to the emerging literature on Thurstonian IRT modeling for implicit motive measures and show that the DTM can be used to study implicit motives across cultures. We describe how a frequently used method to assess DIF in IRT models—the likelihood-ratio method—can be used to analyze and identify DIF in DTM models for implicit motive measures. We also describe how the results of a likelihood-ratio analysis can be used to build a partially invariant DTM model that accounts for pictures with DIF in estimating latent motive scores.

The second contribution is a demonstration of DTM-based cross-cultural analyses using a data set of the Operant Motive Test (OMT; Kuhl & Scheffer, 1999), a measure for implicit motives. This data set includes participants from three cultures: Cameroon, Germany, and Costa Rica (Chasiotis, Hofer, & Campos, 2006). Our demonstration shows how DTM-based analyses can provide insights into differences in picture perception and implicit motives across cultures.

Implicit motives

Implicit motives are described as associative networks that connect situational cues with basic affective reactions and implicit

behavioral tendencies. These behavioral tendencies are nonconscious dispositions to seek or avoid certain classes of incentives (McClelland, 1984; McClelland, Atkinson, Clark, & Lowell, 1953). The incentives are typically classified in three categories: need for affiliation, need for achievement, and need for power. The implicit affiliation motive is described as an interest to establish, maintain, or restore positive relationships with others; the implicit achievement motive as an interest to improve one's performance; and the implicit power motive as the interest to impress, influence, and control others (Schultheiss & Brunstein, 2010).

Cultural differences

In the early years of research on implicit motives, McClelland and colleagues studied cultural aspects of the need for achievement (McClelland, 1961; McClelland et al., 1953). In later years, research was typically limited to European-American cultural contexts. Over a decade ago, researchers started to revive cross-cultural studies on implicit motives.

Cross-cultural research on implicit motives has focused on two lines of research. The first line of research focused on cultural differences in implicit motives and on culture-specific differences in behavioral predictions of implicit motives. Contributing to this first line of research, researchers found cultural differences in need for achievement (e.g., Cameroon vs. Costa Rica: $d = 0.88$; Busch, Hofer, Chasiotis, & Campos, 2013) and need for power (e.g., Germany vs. Cameroon: $d = .58$; Hofer, Busch, Chasiotis, Kärtner, & Campos, 2008). A sample including managers from 24 countries also provided evidence for systematic differences in need for achievement and need for affiliation (van Emmerik, Gardner, Wendt, & Fischer, 2010).

The second line of research focused on universal characteristics and showed similarities in the development and behavioral predictions of implicit motives among people with various cultural backgrounds. Studies contributing to this second line of research have, for instance, shown that a high congruency between implicit and explicit achievement and affiliation motives was related to higher life satisfaction of participants recruited in European-American, Latin American, and African cultures (Hofer & Chasiotis, 2003; Hofer, Chasiotis, & Campos, 2006). Research in this area has also documented that people—across various cultures—who are capable of accessing their implicit motives have a higher congruency between explicit and implicit motives, and commit stronger to self-congruent goals (Hofer et al., 2010). Finally, cross-cultural research has also found support for the universality of the developmental link between a prosocial power orientation and childhood context variables like parental socioeconomic status and number of siblings (Chasiotis, Bender, & Hofer, 2014; Chasiotis et al., 2006). Although both lines of cultural research on implicit motives have made considerable advances, knowledge on the influence of cultural contexts on implicit motives is still limited (Hofer & Chasiotis, 2011).

Bias in cross-cultural research of implicit motive measures

Cross-cultural studies on implicit motive measures come along with various methodological challenges. Van de Vijver (2000) addressed some of those challenges and formulated a

theoretical background to evaluate Thematic Apperception Test (TAT) type measures used in cross-cultural research. The key concept in this evaluation is *bias*, which is defined as “a lack of similarity of psychological meaning of test scores across cultural groups” (Van de Vijver, 2000, p. 88). He described three types of bias: construct bias, method bias, and item bias.

Construct bias exists if the measured construct differs across cultural groups. Sources of construct bias can be only partially overlapping construct definitions or incomplete assessment of various aspects of the construct. *Method bias* occurs when test samples differ in relevant characteristics, when test characteristics such as familiarity with stimulus material differ between countries, or when the person administering the test influences the test score. *Item bias* is present when people with the same level in a certain trait that come from different cultures have a different probability of endorsing an item. Sources for item bias can be that stimulus material, such as a picture, invokes different psychological functions or has a specific meaning only in one of the tested cultures. Thus, item bias is systematic and not the same as noise, or measurement error. Item bias is typically studied using DIF analysis. DIF analysis can be done with IRT and non-IRT techniques.

Researchers have successfully identified DIF pictures in implicit motive tests using non-IRT techniques such as the Mantel–Haenszel procedure (Hofer & Chasiotis, 2004; Hofer, Chasiotis, Friedlmeier, Busch, & Campos, 2005) or log-linear model analysis (Busch et al. 2013; Hofer et al., 2005). In the Mantel–Haenszel procedure, one applies chi-square statistics to test whether the number of motive responses is equally distributed among two cultures for a picture (Hofer & Chasiotis, 2004). In the log-linear model analysis, one tests hierarchically related models that successively include the following effects: score level, culture, and interaction of culture and score level (Hofer et al., 2005). If the model with only score level fits best to the data, an item is considered as not biased. Both procedures categorize respondents score levels into low–medium–high.

In this study, we analyze item bias using an IRT-based approach. The procedure we suggest expands non-IRT techniques and enables precise parameter estimations using a model-based separation of latent motives and item effects. A central feature of the IRT approach is that it enables partial invariance models. In partial invariance models, one estimates item parameters separately for each group in case DIF was detected. Therefore, one does not need to fully exclude DIF items to be able to estimate meaningful group differences in implicit motives.

DIF in implicit motive measures using DTM

DIF is present when an item has a different probability to be solved (or acknowledged) by respondents with the same latent trait who belong to qualitatively different groups, such as groups with different cultural backgrounds. DIF analysis is a statistical method—typically applied in IRT—to detect biased test items. The goal of a DIF analysis is to distinguish between real existing group differences in latent traits and test bias (Thissen et al., 1986).

Dynamic Thurstonian IRT

The most prevalent measurement model for implicit motive tests has always been the sum score model (classical test theory). Because researchers found low internal consistencies (which are part of classical test theory methodology), but also higher retest reliabilities and strong validity evidence, it seems unlikely that the sum score model is adequate (Atkinson & Birch, 1970; Tuerlinckx et al., 2002). A prominent explanation of why classical test theory fails to adequately measure reliability is given in the *dynamics of action* theory (DoA; Atkinson & Birch, 1970). The authors described—among other processes—two response processes that influence the response to stimuli in implicit motive measures. The first process is termed *change of activity* and describes that an individual changes an action when the strength of another action tendency is higher, which implies that action tendencies compete for enactment. The second process is the *consummatory force*, which describes the reduction of an action tendency after enactment. For instance, when a person is hungry, the action strength of eating is high, so that the action of eating will win the competition against other action tendencies. When the person has finished eating, the action tendency for eating will be reduced and other action tendencies will be higher. The typical explanation of this consummatory force is that the need has been satisfied and temporarily loses some of its strength. After a refractory period, the motive strength returns to its original strength. This leads, according to DoA, to a waning and waxing of motive-specific behavior.

With advancements in IRT modeling it became possible to test these processes. Tuerlinckx et al. (2002) tested the assumed reduction of motive strength in a series of models. The first model was a basic IRT model, the second model was a dynamic model that included a reduction of motive strength after enactment, and the third model was a stochastic dropout model that accounts for behavior that was not led by the studied motive. The best fitting model was the third model. These IRT models included only one motive, so that motive competition was not examined.

The most recent and refined measurement model for implicit motives is the DTM model (Lang, 2014) that builds on recent advancements in the modeling of paired comparison data (Böckenholt, 2001) and forced-choice questionnaires (Brown & Maydeu-Olivares, 2013) to also model motive competition. The DTM also models dynamic reduction of motive strength. This model enabled researchers to test the process of reduction in motive strength after enactment suggested in DoA. Studies for the Picture Story Exercise (PSE; Lang, 2014) as well as the OMT (Runge et al., 2016) have used the DTM as a framework to test different conceptualizations of motive reduction after enactment: a temporary effect as suggested in DoA, a longer but still temporary effect that lasts for two or three pictures, as well as a sustained effect that lasts for length of the whole test. These studies showed that a sustained effect fit best to the PSE and OMT data sets that were studied. Lang (2014) speculated that the recovery of implicit motives might take more time than originally anticipated and might happen after respondents finish working on a picture set.

Using DTM for DIF analyses

The advantage of using an IRT framework for studying motive responses is that IRT allows for advanced DIF analysis. The fundamental difference between the previously applied DIF detection procedures and IRT-based DIF detection procedures is that IRT models estimate person and picture effects separately. IRT-based techniques are commonly considered to be less prone to confounding real mean differences with bias (Lord, 1977) and researchers have suggested that IRT methods are “the most generally valid of all biased item detection methods” (Osterlind, 1987, p. 69). Analyzing DIF in implicit motive measures using DTM offers various additional advantages over non-IRT methods.

First, the DTM estimates latent traits based on two underlying response processes (motive competition and dynamic reduction of motive strength) and thus allows for a more precise parameter estimation both for person as well as for picture parameters. Both previous approaches build groups of low–medium–high motive levels (score level) and depend on sum scores. As a consequence, those tests only really function well when most pictures are unbiased (Hofer et al., 2005). On the contrary, IRT-based DIF analysis is able to perform DIF analysis even when only one item is unbiased (Thissen et al., 1986).

Second, IRT DIF analyses are parametric and therefore allow us to compare more than two groups simultaneously, which is especially interesting in studies in more than two cultures. The IRT approach shares this feature with the previously applied log-linear approach. The Mantel-Haenszel procedure can only compare two groups and is, thus, less suitable for cross-cultural research based on three or more cultures.

Third, in the case that DIF is found, it is possible to build a partial invariance model that estimates picture parameters of the DIF categories separately for each group. One does not need to exclude pictures in which DIF was found; thus, more information is included in the trait estimation.

DIF analysis using likelihood-ratio tests

A common strategy to analyze items for DIF in IRT uses likelihood-ratio tests to compare nested models (Thissen et al., 1986). The typical procedure is to compare a basic IRT model in which item parameter estimation is constrained to be equal for all groups with a DIF model that allows the item parameter for the focused item to be estimated separately for the groups of interest (De Boeck et al., 2011; Thissen et al., 1986). If the DIF model improves model fit, the item is flagged as a DIF item.

In this strategy, one assumes that all other items in the DIF model are unbiased, which is in most cases unknown. Because a set of unbiased items is typically unknown, one starts the model comparisons for the first item with the basic IRT model. If the first item shows DIF, all following model comparisons estimate this item separately for each group; if the item shows no DIF, it is estimated equally for all groups. In this procedure, one does not need to make assumptions for already tested items. We then proceed to test all following items in the same way. We also tested a more complex iterative strategy to

establish a set of unbiased items, and results were highly similar.¹ The DTM is a Rasch type model with a fixed a parameter. Thus, only uniform DIF can be tested. However, nonuniform DIF is relatively rare (Dorans & Holland, 1992; Van de Vijver & Leung, 1997).

When DIF has been identified, one can either remove those pictures from the test or specify a partial invariance model. In the partial invariance model, all picture parameters for which DIF has been found are estimated separately for each group and all non-DIF parameters are estimated jointly. This procedure is typically used in structural equation modeling (Reise, Widaman, & Pugh, 1993) and has the advantage that information from the DIF pictures is not lost.

This research

We analyze a data set of an OMT picture set that has been developed and applied for implicit motive research in three different cultures: Cameroon, Germany, and Costa Rica (Chasiotis et al., 2006). We were interested in differences in latent motives between those cultures and OMT pictures that potentially show DIF. We analyze the OMT picture set for DIF in three cultures using the steps previously described and determine unbiased latent mean differences using a partial invariance model.

Method

Selection of cultures

A sample needs to show a high cultural diversity to test universal assumptions. The cultures in this study show high cultural diversity, being recruited from Africa (Cameroon), Europe (Germany), and Latin America (Costa Rica). The cultures in this study differ in socioeconomical background, sociocultural norms, values, and orientations (Hofstede, 2001). The Costa Rican and German samples are relatively representative for the nation, but Cameroon is a multiethnic nation. To control for cultural differences among African participants, the Cameroonian sample is restricted to Nso, which is a large ethnic group in the Anglophone northwest province of Cameroon (Yovsi, 2003).

Participants

The sample of 369 OMTs is part of a research project that aimed at measuring implicit motives in three different cultures (Chasiotis et al., 2006; Hofer et al., 2005) and comprises 180 women and 189 men with an age range from 18 to 75 ($M =$

36.42, $SD = 14.25$) years. The data set includes participants from Cameroon ($n = 125$), Germany ($n = 124$), and Costa Rica ($n = 120$). Participants from Costa Rica and Cameroon were interviewed at their homes, whereas participants from Germany were interviewed at the University of Osnabrück. The sample is balanced for gender, socioeconomic status, and rural versus urban regions within each of the countries. The OMT data from this research project have previously been published in Busch et al. (2013), Chasiotis et al. (2014), Chasiotis et al. (2006), and Hofer et al. (2008).

The Operant Motive Test

The OMT is a recent measure of implicit motives that is based on more traditional picture-based implicit motive test such as PSE and TAT (e.g., Smith, 1992). In the OMT assessment procedure, researchers present a series of pictures depicting social scenes with one or more persons. Respondents are asked to choose one of the persons as main character of their story. The OMT only asks for brief answers to four questions:

1. What is important for the person in this situation and what is the person doing?
2. How does the person feel?
3. Why does the person feel this way?
4. How does the story end?

A distinctive feature is that the OMT uses drawings as picture cues. The drawings are less detailed than photographs and leave certain characteristics such as gender, clothing, and ethnic group open. Neutral drawings reduce the probability that respondents choose a character based on cultural characteristics, which could lead to DIF.

The responses are categorized into one of the three motives, and a story with no motive content is coded as zero. The OMT coding system additionally differentiates among five categories within each motive: three approach, one avoidance, and one in-between category (Kuhl & Scheffer, 2001). Researchers studied the differentiation between approach and avoidance motivation since the early years of research on implicit motives, especially for achievement motivation (McClelland & Liberman, 1949). Different researchers have developed procedures to measure hope of success and fear of failure (Birney, Burdick, & Teevan, 1969; McClelland et al., 1953; Pang, 2006). Veroff and Veroff (1972) suggested that the power motive divides into a hope of power and a fear of weakness component and Boyatzis (1973) suggested that the affiliation motive divides into hope for closeness and fear of rejection. A differentiated review is provided by Schultheiss (2010). The OMT builds on this research and differentiates between approach and avoidance motivation.

To test whether the theoretically based differentiation into approach and avoidance components applies to this data set, we correlated latent approach and avoidance motivations for each motive. A high correlation would indicate that a differentiation between approach and avoidance motivation is not justified for the OMT. No correlation or a negative correlation would indicate that the approach and avoidance components are measuring different aspects. We found negative correlations between approach and avoidance motivation for affiliation ($r = -.35$), achievement ($r = -.37$), and power ($r = -.15$) motivation when approach motivation included from Level 1 to 3. This

¹The iterative procedure to establish a set of unbiased items is performed in three steps (Edelen, Thissen, Teresi, Kleinman, & Ocepek-Welikson, 2006). In a first step, one identifies a set of temporary anchor items. Therefore, one tests every item for DIF under the assumption that all other items are non-DIF. All items that do not show DIF in the first step are then used as potential anchor items; the other items are considered as potential DIF items. In the second step, one establishes the final anchor set of unbiased items. Therefore, one removes all potential DIF items and retests all potential anchor items, under the assumption that the other potential anchor items are unbiased. This step is repeated until no item is identified as a DIF item. The remaining items can then be used as final anchor set. In the final step, each of the potential DIF items identified in Steps 1 and 2 is retested relative to the final set of anchor items. Only items that show DIF in the last steps are considered DIF items.

finding gives evidence that separating approach and avoidance motivation is reasonable for this OMT data set. We used a strict definition of approach motivation and included OMT Levels 1 to 3 into an approach factor, because this model showed higher parameter recovery compared to a broader approach factor that includes Levels 1 to 4 (see later for IRT reliability estimation).²

The OMTs were filled out in three different languages: English in Cameroon, German in Germany, and Spanish in Costa Rica. Four coders at the University of Osnabrück who speak both English and German evaluated the OMTs from Cameroon and Germany. Five Spanish-speaking coders from the University of Costa Rica evaluated the OMTs from Costa Rica. All coders used the detailed information provided in the OMT manual (Kuhl & Scheffer, 2001)—in the Spanish translation for coders from Costa Rica—and were instructed and trained through OMT coding seminars. All coders reached an agreement of Cohen's $\kappa > .85$ on training material prior to the coding of the study. The rater agreement in the study has been evaluated on 20 double-coded and translated OMTs from Costa Rica. Cohen's κ for the 20 translated stories from Costa Rica between the raters from Costa Rica and Germany was .86 for affiliation, .81 for achievement, and .84 for power (Chasiotis & Hofer, 2003). Because German and Costa Rican coders reached sufficient agreement, all following OMTs have been single coded and randomly distributed between all coders. A detailed description of the coding procedure in this research project is given by Chasiotis and Hofer (2003).

We estimated reliability using an IRT simulation approach that describes how well trait estimates are recovered in 100 simulated and bias-corrected simulation runs. A description of this standard procedure for reliability estimation in DTM is provided by Lang (2014). The reliability estimates were $r_{\theta\theta}^2 = .63$ for affiliation, $r_{\theta\theta}^2 = .50$ for achievement, and $r_{\theta\theta}^2 = .69$ for power. The reliabilities for affiliation and power are similar to OMT IRT reliabilities in the literature (Lang et al., 2012; Runge et al., 2016). For research purposes, reliabilities in this range are typically sufficient (Ellis, 2013). The reliability for achievement is lower compared to previous OMT studies.

Analyses

Dynamic Thurstonian IRT model for the OMT

Researchers have recently applied modern Thurstonian IRT models to implicit motive tests and modeled motive competition as choice behavior (Lang et al., 2012) and added the dynamic reduction to the model (Lang, 2014; Runge et al., 2016). In the following section, we describe how the Thurstonian IRT model can be used to analyze OMT data using general linear mixed-effects models.

Motive competition is modeled as pairwise motive comparisons on Level 1. Let $\pi_{affachs}$, $\pi_{affpows}$, $\pi_{affzeros}$, $\pi_{achpows}$, $\pi_{achzeros}$, and $\pi_{powzeros}$ denote the probability that person s prefers one motive category (aff, ach, pow, and zero for other content) to another. These probabilities are a function of difference

between the latent utilities for each motive category. The latent utilities for each motive can be denoted as μ_{affs} , μ_{achs} , μ_{pows} , and μ_{zeros} . They are linked with a probit link to the binomial outcome. The Level 1 probit model can be written as

$$\begin{pmatrix} \text{probit}(\pi_{affachs}) \\ \text{probit}(\pi_{affpows}) \\ \text{probit}(\pi_{affzeros}) \\ \text{probit}(\pi_{achpows}) \\ \text{probit}(\pi_{achzeros}) \\ \text{probit}(\pi_{powzeros}) \end{pmatrix} = \begin{pmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \\ 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu_{affs} \\ \mu_{achs} \\ \mu_{pows} \\ 0 \end{pmatrix} = \mathbf{D}\mu_s$$

In this denotation, μ_s is a vector that contains the latent utilities and \mathbf{D} is the design matrix that contains three dummy variables that refer to three of the four motive categories. The fourth row of dummy variables can be omitted without losing generality, and the equivalent row of mean evaluations is constrained to zero.

In Level 2 of the linear mixed-effects model, the latent utilities μ_s are specified by adding a person parameter, a picture parameter, and additionally a dynamic parameter that accounts for the dynamic motive reduction. The complete Level 2 specifications of the DTM model can be written as follows:

$$\mu_{affs} = \sum_{m=1}^M x_{affms} \beta_{affm} + DE_{affs} \beta_{aff} + v_{affs}$$

$$\mu_{achs} = \sum_{m=1}^M x_{achms} \beta_{achm} + DE_{achs} \beta_{ach} + v_{achs}$$

$$\mu_{pows} = \sum_{m=1}^M x_{powms} \beta_{powm} + DE_{pows} \beta_{pow} + v_{pows}$$

$$\mu_{zeros} = 0.$$

In this specification, v_s is a random person effect (theta) for each motive that captures the latent motive of the person. x_{ms} ($m = 1, \dots, M$) is a picture covariate with M being the total picture number and a fixed effect β_m that denotes the mean evaluation of picture m for the corresponding motive category (in contrast to standard IRT, in which β_m denotes easiness and not difficulty). For each picture, the DTM includes up to three motivational categories (affiliation, achievement, and power) and a zero category that includes all non-motive-relevant responses including all avoidance categories. These response options for each picture are captured by the fixed effect β_m and are conceptually the same as items in other (Thurstonian) IRT models. When we refer to items in the context of this article, we thus refer to the endorsement of a specific motivational response category for a specific picture in the context of the DTM. DE denotes the dynamic effect and is the sum of all previous responses of the respective motive. A detailed model description for the OMT is provided by Runge et al. (2016).

Model estimation

DTM models can be fitted in the lme4 1.1–12 package using the Bayesian glmer function (Bates, Mächler, Bolker, & Walker, 2015) in the R environment (R Core Team, 2015). Some of the

²All analysis can also be applied to the less strict definition of approach using OMT Levels 1 to 4 that also correlate negatively with avoidance motivation (affiliation $r = -.30$, achievement $r = -.47$, and power $r = -.09$).

pictures showed almost no motive responses for one or two motives. Because information in responses with very low response probabilities is limited, it is difficult for generalized linear mixed-effects models (GLMMs) like the DTM to estimate effects. The standard approach is to exclude the item (e.g., Debeer & Janssen, 2013). It is reasonable to assume that some pictures in the OMT do not invoke a certain motive. Thus, we excluded items that show very few responses from model estimation. In detail, items with a response frequency lower than 3% for specific motives in seven cases (Pictures 1–5 and 12; see Table 1) were eliminated before model estimation. The glmer function uses the Laplace approximation (Böckenholt, 2001) and estimates random effects with the maximum a posteriori method (De Boeck et al., 2011). We used the BOBYQA algorithm in NLOPT as an optimizer for model estimation. We reran the analysis with other optimizers and results were highly similar. A detailed description of these procedures including an R code tutorial is provided by Runge et al. (2016).³

Results

Table 1 shows the overall response frequencies for each motive. Table 2 provides the result of the DIF analyses. As indicators of model fit we report the Bayesian information criterion (BIC), as it is common in the literature on generalized linear mixed-effects models (Bates et al., 2015; Doran, Bates, Bliese, & Dowling, 2007). The BIC is an index that is smaller for models that provide better fit to the data and penalizes complex models. With a high sample size, the BIC is consistent in selecting the true model (Vrieze, 2012). Picture 1/affiliation (BIC change: 15.6), Picture 2/affiliation (BIC change: 8.6), and Picture 6/power (BIC change: 3.5) were identified as DIF items. The first DIF item was Picture 1/affiliation that showed a much lower item parameter for Germany compared to Cameroon and Costa Rica (see Table 3). The picture shows two people embracing each other. The second DIF item was Picture 2/affiliation, which also showed a much lower item parameter for Germany compared to Cameroon and Costa Rica. This picture shows two people who speak behind the back of another person. The third DIF item was Picture 6/power, which had a higher item parameter for Costa Rica compared to Cameroon and Germany. This picture shows two people drawing circles.

To evaluate latent mean differences across the tested cultures we analyzed a partial invariance model with item parameters estimated separately for each culture for the three identified DIF items. In this model, we allowed latent mean

Table 1. Frequency of motive responses for each picture.

Picture	Aff %	Ach %	Pow %	Other %
1. Embracing couple	44.7	0.3	23.6	31.4
2. Talking behind the back	7.0	2.4	23.8	66.7
3. Giant in a group	6.0	2.4	44.2	47.4
4. Laying stones	0.0	37.7	0.8	61.5
5. People behind a desk	0.5	7.0	23.8	68.6
6. Drawing circles	1.6	39.8	10.8	47.7
7. Opposite with folded arms	5.1	9.2	36.0	49.6
8. Big and small person	3.5	12.2	10.8	73.4
9. Women sitting behind a man	6.2	3.5	12.8	79.4
10. Group of people	12.7	15.7	27.4	44.2
11. Women and man talking	23.6	3.3	24.1	49.1
12. Man with a fist	0.5	15.7	18.4	65.3

differences in implicit motives to vary between the cultures. The parameter estimates of the final model are presented in Table 3. Higher values for a picture parameter indicate a higher picture pull. The dynamic effects are negative, indicating that a response for a certain motive is less likely for each previous response to this motive. This finding is in line with previous studies (Lang, 2014; Runge et al., 2016).⁴ The information from a DTM can be directly translated to predictions for responses for a particular picture. For instance, consider a hypothetical scenario in which a person with one previous affiliation response (a dynamic effect DE of 1 for affiliation), but no previous achievement responses (a dynamic effect DE of 0 for achievement) and true latent traits of $\theta_{aff} = 2$ and $\theta_{ach} = 1$ for affiliation and achievement, respectively, would answer Picture 7. The predicted probit that this person chooses affiliation over achievement in this picture is

$$\begin{aligned} \pi_{affachs} &= \left(x_{aff7s} \beta_{aff7} + DE_{aff} \beta_{aff} + v_{affs} \right) \\ &\quad - \left(x_{ach7s} \beta_{ach7} + DE_{ach} \beta_{ach} + v_{achs} \right) \\ &= (-1.29 + (1 \times (-1.75)) + 2) \\ &\quad - (-0.57 + (0 \times (-0.79)) + 1) = -1.61. \end{aligned}$$

The probit link in GLMM follows the (inverse) cumulative standard normal distribution. Thus, probits can be translated to a probability that the person chooses affiliation over achievement using a normal distribution table or a statistical software package. A value of -1.61 is equal to a probability of 5%.

Table 4 presents the latent mean differences of implicit motives for each culture from the partial invariance model as well as the original model without any DIF analysis. The values indicate the difference between the cultures in the average response probabilities for each motive. For example, German respondents were least likely and people from Costa Rica were most likely to respond with an affiliation story. People from Germany have a lower latent affiliation motivation than people from Cameroon, and people from Cameroon have a lower latent affiliation motivation than people from Costa Rica.

³The following basic R code is for a DIF model that tests DIF for affiliation in Picture 1: `glmer(response~0+AFF:PIC1:CAMEROON+AFF:PIC1:GERMANY+AFF:PIC1:COSTARICA+AFF:PIC2+AFF:PIC3+AFF:PIC6+AFF:PIC7+AFF:PIC8+AFF:PIC9+AFF:PIC10+AFF:PIC11+ACH:PIC4+ACH:PIC5+ACH:PIC6+ACH:PIC7+ACH:PIC8+ACH:PIC9+ACH:PIC10+ACH:PIC11+ACH:PIC12+POW:PIC1+POW:PIC2+POW:PIC3+POW:PIC5+POW:PIC6+POW:PIC7+POW:PIC8+POW:PIC9+POW:PIC10+POW:PIC11+POW:PIC12+AFF:DEAFF+ACH:DEACH+POW:DEPOW+(AFF+ACH+POW+0|person), family=binomial("probit"), tdata, control=glmerControl(optimizer="nloptwrap", calc.derivs=F))`

⁴We have also tested a model using a temporary reduction in motive strength that is similar to the conceptualization of motive consumption in DoA. We found that, consistent with the literature (Lang, 2014; Runge et al. 2016), the model with temporary reduction in motive strength (BIC = 10241.9) did not improve model fit over the model with sustained reduction in motive strength (BIC = 9778.4).

Table 2. Change in Bayesian information criterion (BIC) in the sequential differential item functioning (DIF) testing.

Model	Aff	Ach	Pow
Picture 1: Embracing couple	15.6	—	−9.5
Picture 2: Talking behind the back	8.6	—	−13.1
Picture 3: Giant in a group	−3.0	—	−4.6
Picture 4: Laying stones	—	−0.7	—
Picture 5: People behind a desk	—	−14.5	−10.9
Picture 6: Drawing circles	−11.2	−11.0	3.5
Picture 7: Opposite with folded arms	−18.3	−17.7	−17.4
Picture 8: Big and small person	−19.6	−15.8	−11.3
Picture 9: Women sitting behind a man	−14.6	−16.4	−11.7
Picture 10: Group of people	−15.2	−18.1	−13.3
Picture 11: Women and man talking	−15.4	−14.2	−13.0
Picture 12: Man with a fist	—	−10.7	−18.0

Note. Bold BIC indicates DIF.

Considerable differences between the normal and the DIF corrected model were found for affiliation motivation in the German part of the sample. The difference reflects the corrected item easiness parameters in the two identified DIF items for affiliation.

Discussion

In this study, we showed how the DTM model can be used to study implicit motives across cultures and identify DIF items in implicit motive measures. We demonstrated the DIF model using a picture set of the OMT. Three items were flagged as DIF. Parameter estimates for these items showed that the first two pictures elicited fewer affiliation responses in the German subsample compared to the other countries and Picture 6 elicited more power responses for Costa Rica compared to the other countries. We built a partial

invariance model in which item parameters were estimated separately for each country for the DIF items. Latent mean differences showed that implicit motives among between the countries. A comparison with the latent mean differences obtained from the standard model showed that the partial invariance model corrected for the DIF items. However, the rank order for latent mean implicit motive scores between the countries did not change and the DIF correction was not large.

The first implication of the study is that this picture set of the OMT is useful for the comparison of implicit motives in the three tested cultures. Although three items were identified as DIF, pictures in this OMT set were mostly unbiased. This picture set can be used for future studies if researchers correct for the identified DIF items or exclude those pictures from model estimations. On a more general level, this study emphasizes earlier research on implicit motive tests that demonstrate that implicit motive tests can be used to compare implicit motives universally across cultures (Hofer et al., 2005).

The second implication of the study is that implicit motives vary between people from different countries with different cultural backgrounds. Our findings suggest that people from Germany had the lowest motive scores among the three groups for all motives. People from Costa Rica had higher affiliation scores than people from Cameroon. This finding is in line with previous research that found differences in implicit motives between people with various cultural backgrounds (Busch et al., 2013; Hofer et al., 2008; van Emmerik et al., 2010) and, thus, indicates the existence of culture-specific variations in implicit motives. Germany having a lower power motivation compared to Costa Rica and Cameroon has previously been reported (Chasiotis

Table 3. Parameter estimates for the partial invariance model.

Estimate	Aff		Ach		Pow	
	Fixed effects	SE	Fixed effects	SE	Fixed effects	SE
Picture 1: Embracing couple					−0.10	0.14
Cameroon	0.24	0.27				
Germany	−0.71	0.30				
Costa Rica	0.69	0.29				
Picture 2: Talking behind the back					−0.49	0.14
Cameroon	−1.32	0.29				
Germany	−3.68	0.60				
Costa Rica	−1.89	0.33				
Picture 3: Giant in a group					0.47	0.14
Picture 4: Laying stones			−0.26	0.15		
Picture 5: People behind a desk			−1.55	0.17	−0.03	0.14
Picture 6: Drawing circles	−2.23	0.31	0.37	0.15		
Cameroon					−0.45	0.21
Germany					−0.58	0.22
Costa Rica					0.34	0.19
Picture 7: Opposite with folded arms	−1.29	0.26	−0.57	0.17	0.76	0.15
Picture 8: Big and small person	−1.64	0.27	−0.47	0.18	−0.10	0.17
Picture 9: Women sitting behind a man	−1.13	0.26	−1.21	0.20	−0.09	0.17
Picture 10: Group of people	−0.01	0.25	0.23	0.18	1.03	0.17
Picture 11: Women and man talking	0.86	0.26	−0.74	0.22	1.06	0.18
Picture 12: Man with a fist			0.18	0.20	0.85	0.19
DE	−1.75	0.13	−0.79	0.10	−0.61	0.05
Random effects		2.20		1.30		1.14
Correlations						
Aff	—					
Ach	0.17		—			
Pow	0.03		0.06		—	

Note. $k = 10,627$ pair-wise comparisons nested in $n = 369$ persons and 12 pictures. Aff = affiliation vs other; Ach = achievement versus other; Pow = power versus other; DE = dynamic effect operationalized as the number of previous motive related responses. SEs are obtained from the summary output in glmer.

Table 4. Differences in latent motives by country.

Picture	Latent mean differences							
	Standard DTM				Partial invariance DTM			
	Germany	SE	Costa Rica	SE	Germany	SE	Costa Rica	SE
Affiliation	−1.26	0.33	0.60	0.30	−1.06	0.34	0.51	0.32
Achievement	−0.84	0.19	0.18	0.18	−0.85	0.19	0.18	0.18
Power	−0.56	0.16	−0.17	0.16	−0.57	0.16	−0.26	0.16

Note. The latent mean differences are centered around Cameroon, which has 0.00 as latent mean motive score for all three motives. DTM = dynamic Thurstonian item-response theory model. SEs are obtained from the summary output in glmer.

et al., 2006). More recent studies (Aydinli, Bender, Chasiotis, van de Vijver, & Cemalcilar, 2015; Chasiotis et al., 2014) replicated the findings—including more than the three countries analyzed in this study—that two factors explain the cultural variance in the development of implicit power motivation: the number of siblings and parental socioeconomic status. Participants with more siblings have a developmental context in which they are exposed to more power-motivated behavior such as caretaking compared to participants with fewer or no siblings. Busch et al. (2013) found, focusing on the flow aspect of achievement, that Germany scored lower, compared to Costa Rica and Cameroon. Achievement motivation can be seen as a materialistic value that is less important in postindustrial cultures like Germany, compared to self-expression and subjective well-being (Inglehart & Baker, 2000). For affiliation motivation, a previous study showed that Costa Rica has a higher need for affiliation compared to Germany; however, this study did not find a lower affiliation motivation for Germany compared to Cameroon (Hofer et al., 2006). Overall, our findings are largely in line with previous studies.

It is difficult to interpret the DIF items because the DIF test does not reveal the source of DIF. One possible explanation for DIF in this study is a systematic bias for coders between Costa Rica and Germany. However, such a bias could only apply to the DIF found in Picture 6, which was between Costa Rica and the other cultures. Although a coder bias could be responsible for the differences in Picture 6, we cannot test this interpretation because the cultures and coders are confounded. None of the other pictures showed DIF between the coders so a general coder bias is unlikely. Another potential post-hoc explanation for the DIF in Pictures 1 and 2 is that Germany is an individualistic country, whereas Cameroon and Costa Rica are more collectivistic countries (Hofstede, 2001). It could be that Pictures 1 and 2 are interpreted as more individualistic in Germany compared to Cameroon and Costa Rica, independent of individual motive strength.

The third implication of the study relates to the scope of the presented DIF detection procedure. This study demonstrated the use of IRT methods in studying DIF for cultures in the OMT. The usefulness of the presented method is, however, not limited to the OMT or cultural research. Another commonly used implicit motive measure is the PSE. Because DTM was developed for the PSE (Lang, 2014) it is straightforward to adapt the presented DIF detection methods for the use in PSE research. Researchers have also been interested in gender (Drescher & Schultheiss, 2016) and age differences (Denzinger, Backes, Job, & Brandstätter, 2016) in

implicit motives. The suggested method can be used to study DIF for those comparisons as well.

Strength and limitations

A strength of our study is that it provides additional evidence for the DTM approach (Lang, 2014) as a psychometric theory of the response processes in implicit motive measures. Psychometricians have repeatedly argued that a latent variable model that provides a theory of the response process in a test is a desirable type of evidence for its validity. Latent variable modeling requires a philosophical position that has been described as entity realism. In entity realism, a causal interpretation may be formulated for the relation between latent variables and their indicators (Borsboom, 2006; Borsboom, Mellenbergh, & van Heerden, 2003, 2004; Edwards & Bagozzi, 2000; Glymour, 2001; Pearl, 2000). This causal relation is commonly restricted to a between-subjects form for latent variable models and can under normal conditions not be generalized to processes at the individual level. Nevertheless, Borsboom et al. (2003) encouraged researchers to build integrative models that incorporate both individual differences and intraindividual variation in items into joint psychometric models. As a first step in this direction, the DTM includes individual dynamic processes and demonstrates that accounting for these processes makes the extraction of meaningful latent between-person variables possible. Future research on the OMT could seek to also include experimental picture manipulations to gain further insights. Efforts of this type would build models similar to the model that Embretson (1994) described for intelligence items. Borsboom et al. (2004, p. 1068) mentioned this model as an important example for a psychometric model that integrates both latent variables and intraindividual cognitive processes and thus bridges the correlational and experimental research in the way they suggested. Research of this type would build on earlier experimental research in the motivation domain that has so far only been conducted using PSE but not OMT measures (e.g., McClelland et al., 1953; McClelland & Liberman, 1949). Testing interactions between motives and experimental conditions would also address Eysenck's and Revelle's call for experimental personality research (Eysenck & Eysenck, 1985; Revelle & Condon, 2015).

A limitation of our demonstration using the OMT is that the OMT is an implicit motive test that is less frequently used than the PSE. Although the OMT likely measures implicit motives, the implicit motive definition of the OMT is not necessarily highly correlated with the motive definition in the Winter coding manual

(Winter, 1994). One study found that only power motivation correlated significantly, but lower ($r = .15$), between OMT and PSE scores. Another study found modest positive correlations of PSE and OMT scores (affiliation $r = .31$, achievement $r = .47$, and power $r = .47$) that were measured 3 months apart (Scheffer, 2001). A strength of the PSE is that some of the pictures have been developed in the context of laboratory experiments.

Another limitation of this study is that the IRT reliability of the achievement motivation measured with this picture set is relatively low. One potential reason could be that the picture set uses fewer pictures (12) compared to the standard picture set of the OMT (15) and the first three pictures show almost no achievement responses, thus the effective test length for the achievement motive is reduced to nine pictures. Researchers recommend a minimum test length of four to five pictures for the PSE (Pang, 2010; Schultheiss & Pang, 2007). The OMT uses more pictures but responses are brief and coders assign only one motive for a picture. Thus, the PSE recommendation is not one-to-one applicable to the OMT. Reliability is also not a linear function of test length; other factors such as cue ambiguity, variance, and dynamic effects can play a role. Thus, it is difficult to give general advice on minimum picture number in the OMT based on this study. Future studies should systematically evaluate the influence of test length on reliability and validity. Researchers have noted that implicit motive measures with a high number of pictures lead to a decrease in reliability and validity as a result of fatigue (Reitman & Atkinson, 1958) or motive consumption (Lang, 2014).

Another limitation is that other response processes that have been suggested are not modeled with the DTM. For instance, Schultheiss, Liening, and Schad (2008) suggested that situations that arouse a specific motive response might differ for individuals, based on differential learning histories. Another process that is not modeled in this study is joint motive expression. A response to a picture could be influenced by multiple motives. For instance, helping behavior can satisfy the need for power as well as the need for affiliation. It is generally possible to model several codings in a story using the DTM (see Lang, 2014), but in the OMT the response questions are designed for brief answers and coders are instructed to only code one motive in each picture on the basis of the idea that most responses are dominated by one motive theme (also see Tuerlinckx et al., 2002). In future studies, researchers could modify the response and coding format of the OMT and model multiple motive expressions in the OMT using DTM. Statisticians have argued that all models are wrong, but some models are useful (Box, 1976). The DTM is a useful model to test response processes in implicit motive measures. Future research might continue this work and model other response processes.

Conclusion

In this study, we showed how the DTM can be used to identify item bias in implicit motive measures and suggested the use of a partial invariance model that corrects for bias, to study implicit motives across culture. We demonstrated the procedure on an OMT data

set with participants from Cameroon, Costa Rica, and Germany. This study further contributes to the emerging IRT literature on implicit motives by displaying an additional application.

References

- Atkinson, J. W., & Birch, D. (1970). *Dynamics of action*. New York, NY: Wiley.
- Aydinli, A., Bender, M., Chasiotis, A., van de Vijver, F. J. R., & Cemalcilar, Z. (2015). Implicit and explicit prosocial motivation as antecedents of volunteering: The moderating role of parenthood. *Personality and Individual Differences*, *74*, 127–132. doi:10.1016/j.paid.2014.10.011
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). doi:10.18637/jss.v067.i01
- Birney, R. C., Burdick, H., & Teevan, R. C. (1969). *Fear of failure*. New York, NY: Van Nostrand-Reinhold.
- Böckenholt, U. (2001). Hierarchical modeling of paired comparison data. *Psychological Methods*, *6*, 49–66. doi:10.1037/1082-989X.6.1.49
- Borsboom, D. (2006). The attack of the psychometricians. *Psychometrika*, *71*, 425–440. doi:10.1007/s11336-006-1447-6
- Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review*, *110*, 203–219. doi:10.1037/0033-295X.110.2.203
- Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review*, *111*, 1061–1071. doi:10.1037/0033-295X.111.4.1061
- Box, G. E. P. (1976). Science and statistics. *Journal of the American Statistical Association*, *71*, 791–799. doi:10.1080/01621459.1976.10480949
- Boyatzis, R. E. (1973). Affiliation motivation. In D. C. McClelland, & R. S. Steele (Eds.), *Human motivation: A book of readings* (pp. 252–276). Morristown, NJ: General Learning Corporation.
- Brown, A., & Maydeu-Olivares, A. (2013). How IRT can solve problems of ipsative data in forced-choice questionnaires. *Psychological Methods*, *18*, 36–52. doi:10.1037/a0030641
- Busch, H., Hofer, J., Chasiotis, A., & Campos, D. (2013). The achievement flow motive as an element of the autotelic personality: Predicting educational attainment in three cultures. *European Journal of Psychology of Education*, *28*, 239–254. doi:10.1007/s10212-012-0112-y
- Chasiotis, A., Bender, M., & Hofer, J. (2014). Childhood context explains cultural variance in implicit parenting motivation: Results from two studies with six samples from Cameroon, Costa Rica, Germany, and PR China. *Evolutionary Psychology*, *12*, 295–317. doi:10.1177/147470491401200203
- Chasiotis, A., & Hofer, J. (2003). *Abschlussbericht für die Deutsche Forschungsgemeinschaft (DFG) "Die Messung impliziter Motive in Deutschland, Costa Rica und Kamerun—Methodische Voruntersuchungen zu einer kulturinformierten Motivationspsychologie"* [Final report of the DFG Project "The measurement of implicit motives in Germany, Costa Rica, and Cameroon—Methodological preliminary investigation for a culturally informed motivational psychology"]. Osnabrück, Germany.
- Chasiotis, A., Hofer, J., & Campos, D. (2006). When does liking children lead to parenthood? Younger siblings, implicit prosocial power motivation, and explicit love for children predict parenthood across cultures. *Journal of Cultural and Evolutionary Psychology*, *4*, 95–123. doi:10.1556/JCEP.4.2006.2.2
- Debeer, D., & Janssen, R. (2013). Modeling item–position effects within an IRT framework. *Journal of Educational Measurement*, *50*, 164–185. doi:10.1111/jedm.12009
- De Boeck, P., Bakker, M., Zwitser, R., Nivard, M., Hofman, A., Tuerlinckx, F., & Partchev, I. (2011). The estimation of item response models with the lmer function from the lme4 package in R. *Journal of Statistical Software*, *39*, 1–28. doi:10.18637/jss.v039.i12
- Denzinger, F., Backes, S., Job, V., & Brandstätter, V. (2016). Age and gender differences in implicit motives. *Journal of Research in Personality*, *65*, 52–61. doi:10.1016/j.jrp.2016.09.003

- Doran, H., Bates, D., Bliese, P., & Dowling, M. (2007). Estimating the multilevel Rasch model: With the lme4 package. *Journal of Statistical Software*, 20, 1–18. doi:10.18637/jss.v020.i02
- Dorans, N. J., & Holland, P. W. (1992). *DIF detection and description: Mantel-Haenszel and standardization*. ETS Research Report Series, 1992. Educational Testing Service, Princeton, NJ. doi:10.1002/j.2333-8504.1992.tb01440.x
- Drescher, A., & Schultheiss, O. C. (2016). Meta-analytic evidence for higher implicit affiliation and intimacy motivation scores in women, compared to men. *Journal of Research in Personality*, 64, 1–10. doi:10.1016/j.jrp.2016.06.019
- Edelen, M. O., Thissen, D., Teresi, J. A., Kleinman, M., & Ocepek-Welickson, K. (2006). Identification of differential item functioning using item response theory and the likelihood-based model comparison approach: Application to the Mini-Mental State Examination. *Medical Care*, 44, S134–S142. doi:10.1097/01.mlr.0000245251.83359.8c
- Edwards, J. R., & Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods*, 5, 155–174. doi:10.1037/1082-989X.5.2.155
- Ellis, J. L. (2013). A standard for test reliability in group research. *Behavior Research Methods*, 45, 16–24. doi:10.3758/s13428-012-0223-z
- Embretson, S. E. (1994). Applications of cognitive design systems to test development. In C. R. Reynolds (Eds.), *Cognitive assessment: A multidisciplinary perspective* (pp. 107–135). New York, NY: Plenum.
- Eysenck, H. J., & Eysenck, M. W. (1985). *Personality and individual differences: A natural science approach*. New York, NY: Plenum.
- Glymour, C. N. (2001). *The mind's arrows: Bayes nets and graphical causal models in psychology*. Cambridge, MA: MIT Press.
- Hofer, J., & Bond, M. H. (2008). Do implicit motives add to our understanding of psychological and behavioral outcomes within and across cultures? In R. M. Sorrentino, & S. Yamaguchi (Eds.), *The handbook of motivation and cognition within and across cultures* (pp. 95–118). Amsterdam, The Netherlands: Elsevier/Academic Press.
- Hofer, J., Busch, H., Bond, M. H., Kärtner, J., Kiessling, F., & Law, R. (2010). Is self-determined functioning a universal prerequisite for motive-goal congruence? Examining the domain of achievement in three cultures. *Journal of Personality*, 78, 747–780. doi:10.1111/j.1467-6494.2010.00632.x
- Hofer, J., Busch, H., Chasiotis, A., Kärtner, J., & Campos, D. (2008). Concern for generativity and its relation to implicit pro-social power motivation, generative goals, and satisfaction with life: A cross-cultural investigation. *Journal of Personality*, 76, 1–30. doi:10.1111/j.1467-6494.2007.00478.x
- Hofer, J., & Chasiotis, A. (2003). Congruence of life goals and implicit motives as predictors of life satisfaction: Cross-cultural implications of a study of Zambian male adolescents. *Motivation and Emotion*, 27, 251–272. doi:10.1023/A:1025011815778
- Hofer, J., & Chasiotis, A. (2004). Methodological considerations of applying a TAT-type picture-story test in cross-cultural research: A comparison of German and Zambian adolescents. *Journal of Cross-Cultural Psychology*, 35, 224–241. doi:10.1177/0022022103262246
- Hofer, J., & Chasiotis, A. (2011). Implicit motives across cultures. *Online Readings in Psychology and Culture*, 4, 5. doi:10.9707/2307-0919.1097
- Hofer, J., Chasiotis, A., & Campos, D. (2006). Congruence between social values and implicit motives: Effects on life satisfaction across three cultures. *European Journal of Personality*, 20, 305–324. doi:10.1002/per.590
- Hofer, J., Chasiotis, A., Friedlmeier, W., Busch, H., & Campos, D. (2005). The measurement of implicit motives in three cultures: Power and affiliation in Cameroon, Costa Rica, and Germany. *Journal of Cross-Cultural Psychology*, 36, 689–716. doi:10.1177/0022022105280510
- Hofer, J., Kärtner, J., Chasiotis, A., Busch, H., & Kiessling, F. (2007). Socio-cultural aspects of identity formation: The relationship between commitment and well-being in student samples from Cameroon and Germany. *Identity: An International Journal of Theory and Research*, 7, 265–288. doi:10.1080/15283480701600744
- Hofstede, G. H. (2001). *Culture's consequences: Comparing values, behaviors, institutions, and organizations across cultures*. Thousand Oaks, CA: Sage.
- Inglehart, R., & Baker, W. E. (2000). Modernization, cultural change, and the persistence of traditional values. *American Sociological Review*, 65, 19. doi:10.2307/2657288
- Kuhl, J., & Scheffer, D. (1999). *Der operante Multi-Motiv-Test (OMT): Manual* [The Operant Multi-motive Test (OMT): Manual]. Osnabrück, Germany: University of Osnabrück.
- Lang, J. W. B. (2014). A dynamic Thurstonian item response theory of motive expression in the picture story exercise: Solving the internal consistency paradox of the PSE. *Psychological Review*, 121, 481–500. doi:10.1037/a0037011
- Lang, J. W. B., Zettler, I., Ewen, C., & Hülshager, U. R. (2012). Implicit motives, explicit traits, and task and contextual performance at work. *Journal of Applied Psychology*, 97, 1201–1217. doi:10.1037/a0029556
- Lord, F. M. (1977). A study of item bias using item characteristic curve theory. In Y. H. Poortinga (Eds.), *Bias problems in cross-cultural psychology* (pp. 19–29). Amsterdam, The Netherlands: Swets & Zeitlinger.
- McClelland, D. C. (1961). *The achieving society*. Princeton, NJ: Van Nostrand.
- McClelland, D. C. (1984). *Motives, personality, and society: Selected papers*. New York, NY: Praeger.
- McClelland, D. C., Atkinson, J. W., Clark, R. A., & Lowell, E. L. (1953). The measuring instrument. In *The achievement motive* (pp. 185–217). East Norwalk, CT: Appleton-Century-Crofts.
- McClelland, D. C., & Liberman, A. M. (1949). The effect of need for achievement on need-related words. *Journal of Personality*, 18, 236–251. doi:10.1111/j.1467-6494.1949.tb01243.x
- Osterlind, S. J. (1987). *Test item bias: Quantitative applications in social sciences*. Beverly Hills, CA: Sage.
- Pang, J. S. (2006). *A revised content-coding measure for hope of success (HS) and fear of failure (FF)*. Unpublished doctoral dissertation, University of Michigan, Ann Arbor, MI.
- Pang, J. S. (2010). Content coding methods in implicit motive assessment: Standards of measurement and best practices for the picture story exercise. In O. C. Schultheiss, & J. C. Brunstein (Eds.), *Implicit motives* (pp. 119–155). New York, NY: Oxford University Press.
- Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge, UK: Cambridge University Press.
- R Core Team. (2015). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. doi:10.1007/978-3-540-74686-7
- Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin*, 114, 552–566. doi:10.1037/0033-2909.114.3.552
- Reitman, W. R., & Atkinson, J. W. (1958). Some methodological problems in the use of thematic apperceptive measures of human motives. In J. W. Atkinson (Eds.), *Motives in fantasy, action, and society* (pp. 664–683). Princeton, NJ: Van Nostrand.
- Revelle, W., & Condon, D. M. (2015). A model for personality at three levels. *Journal of Research in Personality*, 56, 70–81. doi:10.1016/j.jrp.2014.12.006
- Runge, J. M., Lang, J. W. B., Engeser, S., Schüler, J., den Hartog, S. C., & Zettler, I. (2016). Modeling motive activation in the operant motives test: A psychometric analysis using dynamic Thurstonian item response theory. *Motivation Science*, 2, 268–286. doi:10.1037/mot0000041
- Scheffer, D. (2001). *Entwicklungsbedingungen impliziter Motive: Bindung, Leistung & Macht* [Developmental Conditions of Implicit Motives: Affiliation, Achievement & Power]. Osnabrück, Germany: University of Osnabrück.
- Schultheiss, O. C. (2010). Implicit motives. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (pp. 603–633). New York, NY: Guilford.
- Schultheiss, O. C., & Brunstein, J. (2010). Implicit motives. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (3rd ed., pp. 603–633). New York, NY: Oxford University Press.
- Schultheiss, O. C., Lienesch, S. H., & Schad, D. (2008). The reliability of a Picture Story Exercise measure of implicit motives: Estimates of internal consistency, retest reliability, and ipsative stability. *Journal of Research in Personality*, 42, 1560–1571. doi:10.1016/j.jrp.2008.07.008
- Schultheiss, O. C., & Pang, J. S. (2007). Measuring implicit motives. In R. W. Robins, R. C. Fraley, & R. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 322–344). New York, NY: Guilford.

- Smith, C. P. (1992). *Motivation and personality: Handbook of thematic content analysis*. New York, NY: Cambridge University Press.
- Thissen, D., Steinberg, L., & Gerrard, M. (1986). Beyond group-mean differences: The concept of item bias. *Psychological Bulletin*, *99*, 118. doi:10.1037/0033-2909.99.1.118
- Tuerlinckx, F., De Boeck, P., & Lens, W. (2002). Measuring needs with the Thematic Apperception Test: A psychometric study. *Journal of Personality and Social Psychology*, *82*, 448–461. doi:10.1037/0022-3514.82.3.448
- Van de Vijver, F. (2000). The nature of bias. In *Handbook of cross-cultural and multicultural personality assessment; Personality and clinical psychology series*. (pp. 87–106).
- Van de Vijver, F. J. R., & Leung, K. (1997). *Methods and data analysis for cross-cultural research* (Vol. 1). Thousand Oaks, CA: Sage.
- van Emmerik, H., Gardner, W. L., Wendt, H., & Fischer, D. (2010). Associations of culture and personality with McClelland's motives: A cross-cultural study of managers in 24 countries. *Group & Organization Management*, *35*, 329–367. doi:10.1177/1059601110370782
- Verhelst, N. D., & Glas, C. A. W. (1993). A dynamic generalization of the Rasch model. *Psychometrika*, *58*, 395–414. doi:10.1007/BF02294648
- Veroff, J., & Veroff, J. B. (1972). Reconsideration of a measure of power motivation. *Psychological Bulletin*, *78*, 279–291. doi:10.1037/h0033184
- Vrieze, S. I. (2012). Model selection and psychological theory: A discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychological Methods*, *17*, 228. doi:10.1037/a0027127
- Winter, D. G. (1994). *Manual for scoring motive imagery in running text*. Ann Arbor, MI: University of Michigan.
- Yovsi, R. D. (2003). *Ethnotheories about breastfeeding and mother–infant interaction*. New Brunswick, NJ: Transaction.