Testing the Unidimensionality of Items

2 authors:

Matthias Ziegler
Humboldt-Universität zu Berlin
259 PUBLICATIONS   **7,730** CITATIONS

SEE PROFILE

Dirk Hagemann
Universität Heidelberg
123 PUBLICATIONS   **5,495** CITATIONS

SEE PROFILE

Editorial

Testing the Unidimensionality of Items

Pitfalls and Loopholes

Matthias Ziegler[1] and Dirk Hagemann[2]

[1]Humboldt-Universität zu Berlin, Germany, [2]Universität Heidelberg, Germany

Constructing a test, an interview, or a behavior observation often aims at deriving scores that are interpreted as manifestations of a trait. In order to test whether such an interpretation is feasible, different psychometric criteria have been suggested (Campbell & Fiske, 1959; Cronbach, 1947; Cronbach & Meehl, 1955; Loevinger, 1957). In recent editorials we outlined the importance of criteria such as construct validity (Schweizer, 2012; Ziegler, Booth, & Bensch, 2013) or setting up the actual evaluation process along the lines of the ABC of test construction (Ziegler, 2014b). Within this editorial we want to put the spotlight onto an often neglected aspect, relevant to the interpretation of test scores – the dimensionality of the underlying items.

are considered to be manifestations of latent variables. In the same sense, test scores are seen as reflections of the standing on the latent construct. Depending on the dimensionality of the items, the test score could be multidimensional. To complicate things, unidimensionality of the test score does not necessarily imply that the items have to measure only one psychological process (Bejar, 1983). If we assume that differences within a latent construct are due to a set of different psychological processes, the test score should adequately reflect those processes. If all items measure the same processes to the same extent, they can still be considered unidimensional (Fischer, 1997).

Defining Unidimensionality

The term unidimensionality is often used in publications in order to describe items or test scores. Within the editorial, both uses are of importance and we want to quickly introduce them. Many psychological measures are constructed in order to assess latent constructs. The differences observed in the answers to the items are considered manifestations of the differences on the latent construct (Cronbach & Meehl, 1955) plus measurement error. An item is considered unidimensional if the systematic differences within the item variance are only due to one variance source, that is, one latent variable. This idea is used to test the unidimensionality of a set of items using the principle of local independence (Lazarsfeld, 1959). According to this principle, a set of items is seen as unidimensional if there are no correlated residuals between the items once the variance due to the latent construct is controlled for. However, the term unidimensionality can also be applied to describe the dimensionality of test scores. As we said before, items

Importance of Unidimensionality for Psychological Assessment

Items (or interview questions or anchors on a behavior anchored rating scale) are often derived from the definition of the construct intended to be measured. This rational method of item development is meant to ensure that all items really capture the construct aimed at and only this construct. As a consequence, items belonging together in a scale are believed to capture differences in the same underlying construct. From a practical perspective this unidimensionality is extremely important. Let us assume the following diagnostic process as an example. The diagnostic process starts with a general question (e.g., "Does Mr. Z. suffer from a major depressive episode?") which is then worked out into a number of specific hypotheses (e.g., the DSM V criteria for depression, such as: "Depressed mood most of the day, nearly every day, as indicated by either subjective report" . . . or "observation made by others" . . .). In a next step, psychological measures (e.g., Center for Epidemic Studies Depression Scale, CES-D;

Radloff, 1977) are chosen to obtain data which can provide information regarding those hypotheses. Thus, test scores derived from psychological measures are directly linked to constructs named in hypotheses (e.g., depressive mood). The items within the chosen test should now reflect differences in exactly this construct and not any other, however related, constructs (e.g., anxiety). Otherwise, the interpretation of the test score as being representative for the construct mentioned in the hypothesis could be wrong. If an item captured not only depressive mood but also anxiety, the total score would contain this information as well. The test score could now not be interpreted as the standing of the person on the latent variable depressive mood. Therefore, unidimensionality of the items comprising a test score is essential for the soundness of the assessment processes the score is being used in. Without testing for unidimensionality, an interpretation of the test score as representing one dimension is potentially risky. Moreover, testing for unidimensionality of the items provides general information regarding factorial validity of the test score interpretation and might reveal that the test score needs to be separated into several scores (Stout, 1987).

It has to be noted here that other psychometric criteria such as reliability or validity evidence for the test score interpretation are clearly not of less importance (Ziegler, 2014b). However, during the revision processes here at EJPA the issue of the dimensionality of the items has come up again and again, often with unsatisfactory outcomes. Thus, this editorial takes up this important issue.

Unidimensionality and Item Construction

Unidimensionality of the items oftentimes plays an important role, explicitly or implicitly, during the initial item construction process. The question A in the ABC of test construction (Ziegler, 2014) asks for the construct to be measured. The answer to this question should not only be the definition of the actual construct but also an overview of its nomological net. The nomological net directly reflects back onto the item construction process. In order to ensure unidimensionality of the items, they should be positioned in the nomological net to ensure that they belong to the construct intended to be measured but not closely related constructs. This, however, is an extremely difficult if not impossible task. It has long been known (Stout, 1987) that items often do not only capture the intended trait but also other aspects, for example, other traits or specifics of the test person (e.g., measurement time). However, when constructing the items, one can at least try to ensure that the items are not loaded with other traits. Otherwise, constructs with a tightly woven nomological net, that is, many overlapping or closely related other constructs, might be impossible to represent with unidimensional items. The pictures in Figures 1a–1f illustrate three common and problematic cases of multidimensionality.
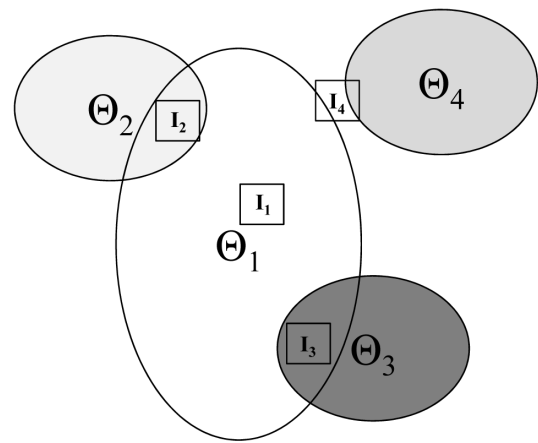

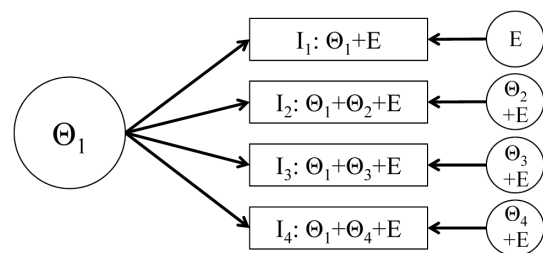*Figure 1a.* Nomological net for Case I.

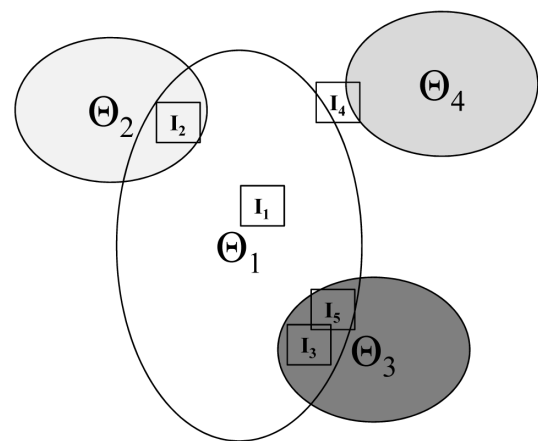*Figure 1b.* Measurement model resulting from the nomological net of Case I.

Figure 1c. Nomological net for Case II.

In Figures 1a, 1c, and 1f, the nomological net for a latent trait Θ_1 is schematically depicted. The net has two other overlapping traits Θ_2 and Θ_3, and one closely related construct Θ_4. The difference between the pictures lies in the position of the items within this net. In Figure 1a, Case I, four items are positioned within the net. Item 1 (I_1) is prototypical for the construct and right at its center. Two of the other items (I_2 and $_3$) can be positioned within the overlapping areas shared with the other traits. The fourth
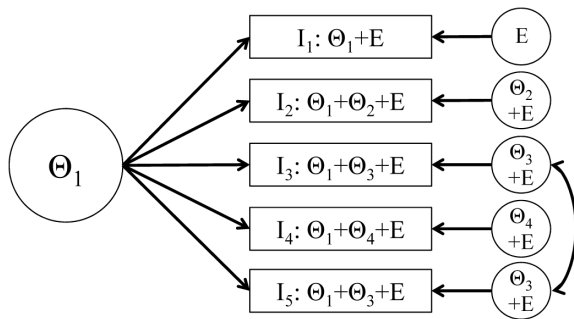
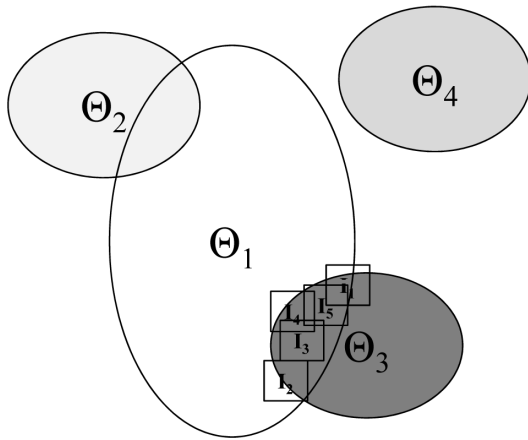*Figure 1d.* Measurement model resulting from the nomological net of Case II.

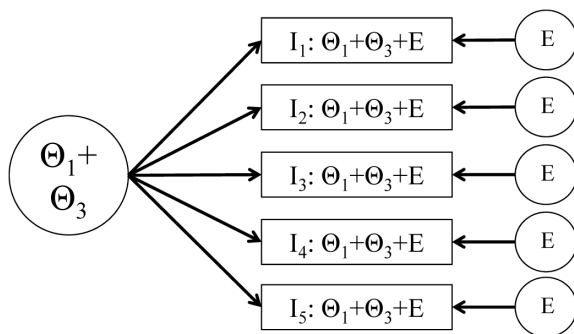*Figure 1e.* Nomological net for Case III.

Figure 1f. Measurement model resulting from the nomological net of Case III.

item links Θ_1 and Θ_4. Figure 1b is an illustration of the measurement model resulting from Case I. It can be seen that the variance shared by all items can be explained by a latent trait. All residuals are uncorrelated and contain measurement error E. For three items the residuals also contain systematic trait variance due to the other traits. Here, the latent variable might be considered unidimensional. Yet, the items are strictly speaking not unidimensional. However, this is not easy to detect, especially if the variance due to the other traits is small.

In Case II (see Figure 1c) an additional item is positioned in the overlap with trait three. The measurement model (see Figure 1d) now contains correlated residuals for items three and five, both of which contain the same systematic variance in their respective residuals. Here it can easily be concluded that the two items sharing correlated residuals are not unidimensional.

Figures 1e and 1f depict Case III which is more problematic. Here, all five items are positioned within the overlap with trait three. The measurement model, however, clearly shows that all items are unidimensional according to the principle of local independence. Yet, when we look at the variance within the latent variable, we see that there actually are two underlying traits. Whereas this could be unproblematic if the two traits represent psychological processes representing the same construct, Case III becomes problematic if this is not true. Figure 1f also shows that the measure built does not capture trait one but rather the specific overlap this trait has with trait three. This case can cause severe problems regarding content validity, especially when item loadings are used to select items (Ziegler, 2014a; Ziegler, 2014b). It is also plausible that a special situation of this case occurs. Here, the items would all capture the trait focused but also a trait like trait 4, that is, a closely related trait. For example, if the items all assess conscientiousness and test takers now fake on the items (Ziegler, 2015; Ziegler, MacCann, & Roberts, 2011), the test score would reflect differences in conscientiousness and faking behavior.

Consequently, the goal of unidimensionality already poses requirements on the item construction process. It seems advantageous to formulate items and then to position them within the nomological net of the construct intended to be measured (Loevinger, 1957). This way, the unidimensionality of the scores and respective items is influenced at this stage of test construction. In later stages of the test construction process, the idea of unidimensionality is often tested thoroughly using a diverse set of analytical tools, for example, exploratory or confirmatory factor analysis or item response theory.

Testing for Unidimensionality – Means and Misunderstandings

Testing the unidimensionality of items has very strong technical parallels to testing the dimensionality of items. Therefore, there are some classical approaches with factor analyses as the most widely used technique. However, there are also several approaches based on an item response technique (IRT) framework.

Exploratory Factor Analysis (EFA)

Before going into the details behind this method, it has to be noted from the start that EFA is not really an appropriate technique to *test* unidimensionality. Rather than that EFA is a viable approach to deriving hypotheses regarding the

possible number of factors underlying the item intercorrelations. Thus, EFA should not be sold in a paper as providing evidence regarding unidimensionality.

When the result of an item construction process is a set of items with no clear idea of the number of factors they constitute, EFA is certainly a good first step. However, EFA requires the test constructor to make a number of subjective decisions. We will focus on three such decisions here: Factor retention rules, methods of factor extraction, and factor rotation.

In order to determine the number of factors retention rules such as parallel analysis (Horn, 1965) or the minimum average partial test (Velicer, 1976) are preferable to methods such as the Scree test or the Eigenvalue larger than 1 rule (Henson & Roberts, 2006). Both factor retention rules are implemented in the R package psych (Revelle, 2014).

Another vital decision regards the method to extract factors. Fabrigar, Wegener, MacCallum, and Strahan (1999) concluded that principal axis factoring should be preferred to principal component analysis when the aim of the research is to identify latent constructs. As this is most often the case for papers submitted to EJPA we adhere to this recommendation.

Finally, there is the question of the rotation method to be used. Rotation methods are used to provide loading pattern matrices that are more easily interpretable because the number and size of the cross-loadings are changed. Generally, there are two approaches. Some rotation methods reduce cross-loadings resulting in larger factor intercorrelations, others produce smaller factor intercorrelations but more cross-loadings (Sass & Schmitt, 2010). In the most extreme form, factor intercorrelations can be set to zero (orthogonal rotation). Schmitt and Sass (2011) demonstrated that the chosen rotation method potentially influences the interpretation of the items' dimensionality. For example, if an orthogonal rotation is used even though the underlying latent constructs are correlated, items might have cross-loadings that are in fact not due to the multidimensionality of the item (as implied by the cross-loadings) but the wrong rotation method. In other words, the test constructor might decide to eliminate an item because it is an impure measure of the construct while in fact this is not true. Schmitt and Sass concluded that there probably is not the best rotation method. Rather than that they recommended to apply different methods to the same data set. If the solutions converged, this could be seen as a robust finding. However, if different methods allowed for strongly differing interpretations, caution and thinking would be warranted instead of adhering to some numerical cutoffs (see also Sass & Schmitt, 2010).

Importantly, EFA results obtained during test construction are often used to select items. To this end, items with substantial cross-loadings are often, and sometimes erroneously, eliminated (Ziegler, 2014a). With regard to unidimensionality, this item selection seems to ensure that no other relevant variance sources are part of the items. However, as we illustrated in Case III in Figures 1a–1f, this can be misleading if all items contain the same two sources of variance. Moreover, as was shown in Case I in Figures 1a–1f, all items loading on the same factor does also not necessarily imply unidimensionality of the items.

Considering this along with the fact that EFA is not a test but rather a hypothesis generating approach, it should only be considered as a starting point to testing dimensionality if no a priori model can be specified.

Confirmatory Factor Analysis (CFA)

Within a CFA the measurement model for each score can be specified and it can be tested whether the data collected with the instrument reflect the theoretical model. If the idea of local independence and thus unidimensionality holds, the measurement model for a test score should not contain correlated errors or loadings from other latent variables. If a test for such a model holds (Heene, Hilbert, Draxler, Ziegler, & Bühner, 2011), we often conclude that the items are unidimensional. This sounds as if the perfect method has been found. In fact, there are numerous papers being submitted to EJPA that present exactly such analyses. As was pointed out before, factorial validity evidence alone oftentimes is not sufficient to warrant a publication. In addition, there are often problems regarding the actual evidence provided.

Again, we can refer to the three cases depicted in Figures 1a–1f. As was the case for EFA, finding fitting models does not necessarily equate to unidimensional items, especially when the test for unidimensionality only includes measurement models alone. Here, only the problem depicted in Case II will reveal itself in the form of a correlated residual. Cases I and III will not result in such correlated residuals. However, as can be seen in Case I, the variance contained in the residuals is not purely unsystematic error. Thus, the dimensionality of the items is not correctly represented in this case. In Case III there is no systematic variance in the residuals. However, the variance captured in the latent variable is not only due to one dimension. Thus, again there is no unidimensionality.

Complicating the picture is the unclear application of rules set to determine model fit (Schweizer, 2010). Oftentimes the rules suggested by Hu and Bentler (1998) are applied. Clearly, those rules have helped to move the field forward in terms of finding general guidelines. However, they have not remained without critique, especially with regard to fitting personality questionnaire data (Marsh, 2004; Marsh et al., 2009; Marsh, 2004). Heene et al. (2011) could show that the conventionally used cutoffs suggested by Hu and Bentler are not necessarily correct when loadings are lower (as is often the case for questionnaires). They concluded that it is necessary to model and discuss misfit before interpreting model parameters. This strategy, however, often goes along with specifying correlated residuals or cross-loadings which as a consequence necessitates explanations and replications. For example, if a few items in a scale all start with "It is important to me...", correlated residuals are likely to occur. In that case they would represent method variance and would not be considered an important violation of unidimensionality. Such straightforward explanations for correlated residuals are not always obvious. Moreover, modeling shared residual variance as

correlations instead of as a latent variable might lead to the underestimation of the actual amount of unwanted additional variance (Lance, Noble, & Scullen, 2002).

Thus, CFA is a step forward compared to EFA when it comes to testing for unidimensionality. However, there are many potential pitfalls waiting along the way. The most common mistakes are the isolated testing of measurement models, the unreflected use of correlated residuals, and the pigheaded use of model fit rules instead of the balanced exploration of misfit sources.

Item Response Technique (IRT)

It is clearly beyond the scope of the present editorial to present a comprehensive overview of IRT, its methods, and their pros and cons. It is important to note though that unidimensionality is one of the main issues in this area (Stout, 2002).

Rost (2001) summarized different IRT models of the Rasch family and their properties. He differentiated the ordinal and dichotomous Rasch models into three subtypes: the mixed Rasch model (Rost, 1991; Rost, Carstensen, & Von Davier, 1997), the linear logistic Rasch model (Fischer, 1997), and the multidimensional Rasch model (McDonald, 1967). The mixed Rasch model can be applied to test person homogeneity, that is, whether all persons use the same trait to answer the items, and is especially useful to detect response biases (Wetzel, Böhnke, Carstensen, Ziegler, & Ostendorf, 2013; Ziegler, 2015; Ziegler, Maaß, Griffith, & Gammon, 2015). The linear logistic Rasch model can be applied when the actual trait intended to be measured consists of several processes. This also seems like a straightforward approach in scenarios such as Case III. It should be noted here that a multi-trait-multi-method SEM can be used for comparable purposes. Importantly especially the family of multidimensional IRT models seems to be suited to test for unidimensionality of the items as discussed here (Hartig & Höhler, 2009). For a more in-depth discussion regarding IRT models and unidimensionality we recommend Reise, Cook, and Moore (2014). Those authors concluded that IRT models are pretty robust, that is, good at detecting multidimensionality in case of multiple latent dimensions or a strong general factor causing additional variance in the items. However, the authors also conclude that IRT models are less robust in other circumstances and warrant caution.

To sum up, none of the presented methods can be seen as a silver bullet which can be used to test unidimensionality. As is often the case, the combination of different methods seems advisable. In the following paragraph we want to outline other potential remedies.

Potential Loopholes

Besides the combination of different test approaches, we want to suggest potential loopholes specifically suited to deal with the cases presented in Figures 1a–1f.

Multi-Trait-[Multi-Method] (MTMM) CFA

The analysis of multi-trait-multi-method data has been proposed by Campbell and Fiske (1959) as a means to provide evidence for convergent and discriminant validity of a test score interpretation. Here, we want to suggest that this approach is a viable solution to the problems presented in Cases I and III. According to the ABC of test construction, defining the construct to be measured and its nomological net also informs the validation strategy. Clearly, the other constructs within the net should be assessed in order to determine discriminant validity. However, assessing those constructs also helps to test unidimensionality. Above we have argued that testing measurement models in isolation does not allow to find the problems depicted in Cases I and III. However, if the other traits in the nomological net are also assessed and modeled in the same structural model as the items of the target measure, the systematic variance within the items accounted for by the overlapping traits will result in modification indices asking for correlated residuals or cross-loadings. Thus, the lacking dimensionality of the items will become apparent and Case I could be detected. Likewise, if in addition to the target measure the overlapping construct is indicated by a separate set of items, a correlation between the latent variables of the target measure and the overlapping construct would reveal Case III and thus help to identify the lack of discriminant validity. Furthermore, by adding different methods to assess the traits, the influence of the specific methods on the findings can be estimated. However, we realize that this is quite a task. At the same time the simultaneous modeling of multiple traits already is a substantial improvement.

Thus, applying a series of tests starting with testing the measurement models for each new scale score (testing for Case II), followed by testing structural models including measurement models for overlapping traits (tested for Cases I and III), should provide much better evidence with regard to unidimensionality.

Importantly, the detection of such multidimensionality does not render the instrument useless. Instead, several approaches are feasible to deal with the issue. Common to all approaches is that they model the different variance sources thereby teasing apart the variance of the items (Brunner & Süß, 2005). With regard to assessment this approach might require the use of factor scores or similar parameters instead of simple sum scores. Otherwise, the different constructs might not be adequately represented. This in turn also necessitates the estimation of reliability using estimators such as McDonald's Omega (Beauducel, 2013; Ziegler & Brunner, in press; Zinbarg et al., 2005).

Experiments

Another method that has great potential for detecting multidimensionality due to overlapping constructs is experimentation. Again, the nomological net must be explicated as a prerequisite, thus the target construct and at least some of the neighboring constructs must be known beforehand. In order to check a potential confounding of the target

measure with undesirable constructs, an experiment may be set up which uses the target measure as a dependent variable and some operationalization of the neighboring constructs as independent variables. An experimental effect of the independent variables on the dependent variables would reveal multidimensionality of the target variable in terms of contamination with undesirable constructs. The empirical effect sizes would increase with the degree of contamination, such that Case I may produce the smallest effect size and Case III may produce the largest. One advantage of this method is that it allows for causal conclusions with regard to the effects of the alleged confounds and one disadvantage of this method is that an operationalization of the neighboring constructs may be difficult to realize.

Editorial Guidelines

At this point we want to put forward a few recommendations that we consider as important essentials in a test evaluation process with regard to testing unidimensionality.

1. According to the ABC of test construction (Ziegler, 2014b) the construct to be measured and its nomological net should be defined. These definitions should be used to position items within the nomological net and to guide the evaluation strategy.
2. Unidimensionality should be tested with a confirmatory framework. CFA or specific IRT models seem most appropriate.
3. If model testing reveals misfit, the source of the misfit should be explored and modeled. Interpretations should pay attention to those misfit sources.
4. A series of model tests (first measurement models then structural models as explained above) should be applied in order to rule out common cases of multidimensionality. As an alternative, a series of experiments may be conducted to identify multidimensionality due to construct overlap.
5. If the items are found to be multidimensional, reliability estimation should be performed using estimators such as McDonald's Omega which allows to model all variance sources and estimate their reliability.
6. Before strong conclusions can be drawn results regarding the model structure should ideally be replicated.

References

Beaducel, A. (2013). Taking the error term of the factor model into account: The factor score predictor interval. *Applied Psychological Measurement, 37*, 289–303. doi: 10.1177/0146621613475358

Bejar, I. I. (1983). *Achievement testing: Recent advances.* Beverly Hills, CA: Sage.

Brunner, M., & Süß, H. M. (2005). Analyzing the reliability of multidimensional measures: An example from intelligence research. *Educational and Psychological Measurement, 65*, 227–240.

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

Cronbach, L. J. (1947). Test "reliability": Its meaning and determination. *Psychometrika, 12*, 1–16.

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

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*, 272–299.

Fischer, G. (1997). Unidimensional linear logistic Rasch models. In W. van der Linden & R. Hambleton (Eds.), *Handbook of Modern Item Response Theory* (pp. 225–243). New York, NY: Springer.

Hartig, J., & Höhler, J. (2009). Multidimensional IRT models for the assessment of competencies. *Studies in Educational Evaluation, 35*, 57–63.

Heene, M., Hilbert, S., Draxler, C., Ziegler, M., & Bühner, M. (2011). Masking misfit in confirmatory factor analysis by increasing unique variances: A cautionary note on the usefulness of cutoff values of fit indices. *Psychological Methods, 16*, 319–336.

Henson, R., & Roberts, J. (2006). Use of exploratory factor analysis in published research: Common errors and some comment on improved practice. *Educational and Psychological Measurement, 66*, 393.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*, 179–185.

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3*, 424–453.

Lance, C. E., Noble, C. L., & Scullen, S. E. (2002). A critique of the correlated trait-correlated method and correlated uniqueness models for multitrait-multimethod data. *Psychological Methods, 7*, 228–244.

Lazarsfeld, P. F. (1959). Latent structure analysis. In S. Koch (Ed.), *Psychology: A study of a science* (Vol. 3, pp. 476–543). New York, NY: McGraw-Hill.

Loevinger, J. (1957). Objective tests as instruments of psychological theory: Monograph Supplement 9. *Psychological Reports, 3*, 635–694.

Marsh, H. W., Hau, K. T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling, 11*, 320–341.

Marsh, H. W., Muthén, B., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A. J. S., & Trautwein, U. (2009). Exploratory structural equation modeling, integrating CFA and EFA: Application to students' evaluations of university teaching. *Structural Equation Modeling: A Multidisciplinary Journal, 16*, 439–476.

Marsh, H. W., Wen, Z., & Hau, K. T. (2004). Structural equation models of latent interactions: Evaluation of alternative estimation strategies and indicator construction. *Psychological Methods, 9*, 275–300.

McDonald, R. P. (1967). Factor interaction in nonlinear factor analysis*. *ETS Research Bulletin Series, 1967*, i-18.

Radloff, L. S. (1977). The CES-D scale a self-report depression scale for research in the general population. *Applied Psychological Measurement, 1*, 385–401.

Reise, S. P., Cook, K. F., & Moore, T. M. (2014). Evaluating the impact of multidimensionality on unidimensional item response theory model parameters. In S. P. Reise & D. A. Revicki (Eds.), *Handbook of Item Response Theory Modeling: Applications to Typical Performance Assessment* (pp. 13–YY). New York, NY: Routledge.

Revelle, W. (2014). *psych: Procedures for Personality and Psychological Research (Version 1.4.5)*. Evanston, IL: Northwestern University.

Rost, J. (1991). A logistic mixture distribution model for polychotomous item responses. *The British Journal of Mathematical and Statistical Psychology, 44*, 75–92.

Rost, J. (2001). The growing family of Rasch models. In A. Boomsma, M. J. van Duijn, & T. B. Snijders (Eds.), *Essays on Item Response Theory* (Vol. 157, pp. 25–42). New York, NY: Springer.

Rost, J., Carstensen, C. H., & Von Davier, M. (1997). Applying the mixed Rasch model to personality questionnaires. In J. In Rost & R. E. Langeheine (Eds.), *Applications of latent trait and latent class models in the social sciences* (pp. XX–YY). New York, NY: Waxmann.

Sass, D. A., & Schmitt, T. A. (2010). A comparative investigation of rotation criteria within exploratory factor analysis. *Multivariate Behavioral Research, 45*, 73–103.

Schmitt, T. A., & Sass, D. A. (2011). Rotation criteria and hypothesis testing for exploratory factor analysis: Implications for factor pattern loadings and interfactor correlations. *Educational and Psychological Measurement, 71*, 95–113.

Schweizer, K. (2010). Some guidelines concerning the modeling of traits and abilities in test construction. *European Journal of Psychological Assessment, 26*, 1–2.

Schweizer, K. (2012). On issues of validity and especially on the misery of convergent validity. *European Journal of Psychological Assessment, 28*, 249–254.

Stout, W. (1987). A nonparametric approach for assessing latent trait unidimensionality. *Psychometrika, 52*, 589–617.

Stout, W. (2002). Psychometrics: From practice to theory and back. *Psychometrika, 67*, 485–518.

Velicer, W. F. (1976). Determining number of components from matrix of partial correlations. *Psychometrika, 41*, 321–327.

Wetzel, E., Böhnke, J. R., Carstensen, C. H., Ziegler, M., & Ostendorf, F. (2013). Do individual response styles matter? *Journal of Individual Differences, 34*, 69–81.

Ziegler, M. (2014a). Comments on item selection procedures. *European Journal of Psychological Assessment, 30*, 1–2.

Ziegler, M. (2014b). Stop and state your intentions!: Let's not forget the ABC of test construction. *European Journal of Psychological Assessment, 30*, 239–242.

Ziegler, M. (2015). "F*** you, I won't do what you told me!" – response biases as threats to psychological assessment. *European Journal of Psychological Assessment, 31*, 153–158.

Ziegler, M., Booth, T., & Bensch, D. (2013). Getting entangled in the nomological net. *European Journal of Psychological Assessment, 29*, 157–161.

Ziegler, M., & Brunner, M. (in press). Test Standards and Psychometric Modeling. In A. A. Lipnevich, F. Preckel, & R. Roberts (Eds.), *Psychosocial Skills and School Systems in the Twenty-First Century: Theory, Research, and Applications*. Göttingen, Germany: Springer.

Ziegler, M., Kemper, C. J., & Kruyen, P. (2014). Short scales – Five misunderstandings and ways to overcome them. *Journal of Individual Differences, 35*, 185–189.

Ziegler, M., Maaß, U., Griffith, R., & Gammon, A. (2015). What is the nature of faking? Modeling distinct response patterns and quantitative differences in faking at the same time. *Organizational Research Methods, 18*, 679–703.

Ziegler, M., MacCann, C., & Roberts, R. D. (2011). Faking: Knowns, unknowns, and points of contention. In M. Ziegler, C. MacCann, & R. R. Roberts (Eds.), *New perspectives on faking in personality assessment* (pp. 3–16). New York, NY: Oxford University Press.

Ziegler, M., Poropat, A., & Mell, J. (2014). Does the length of a questionnaire matter? *Journal of Individual Differences, 35*, 250–261.

Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's α, Revelle's β, and McDonald's ω H: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika, 70*, 123–133.

Matthias Ziegler

Institut für Psychologie
Humboldt Universität zu Berlin
Rudower Chaussee 18
12489 Berlin
Germany
Tel. +49 30 2093-9447
Fax +49 30 2093-9361
E-mail zieglema@hu-berlin.de

Dirk Hagemann

University of Heidelberg
Institute of Psychology
Hauptstrasse 47-51
69117 Heidelberg
Tel. +49 6221 54-7283
Fax +49 6221 54-7325
E-mail dirk.hagemann@psychologie.uni-heidelberg.de