

# Model Specification Searches in Structural Equation Modeling Using Bee Swarm Optimization

Educational and Psychological  
Measurement  
2024, Vol. 84(1) 40–61  
© The Author(s) 2023



Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/00131644231160552  
journals.sagepub.com/home/epm



Ulrich Schroeders<sup>1</sup>, Florian Scharf<sup>1</sup>, Gabriel Olaru<sup>2</sup>

## Abstract

Metaheuristics are optimization algorithms that efficiently solve a variety of complex combinatorial problems. In psychological research, metaheuristics have been applied in short-scale construction and model specification search. In the present study, we propose a bee swarm optimization (BSO) algorithm to explore the structure underlying a psychological measurement instrument. The algorithm assigns items to an unknown number of nested factors in a confirmatory bifactor model, while simultaneously selecting items for the final scale. To achieve this, the algorithm follows the biological template of bees' foraging behavior: Scout bees explore new food sources, whereas onlooker bees search in the vicinity of previously explored, promising food sources. Analogously, scout bees in BSO introduce major changes to a model specification (e.g., adding or removing a specific factor), whereas onlooker bees only make minor changes (e.g., adding an item to a factor or swapping items between specific factors). Through this division of labor in an artificial bee colony, the algorithm aims to strike a balance between two opposing strategies diversification (or exploration) versus intensification (or exploitation). We demonstrate the usefulness of the algorithm to find the underlying structure in two empirical data sets (Holzinger–Swineford and short dark triad questionnaire, SDQ3). Furthermore, we illustrate the influence of relevant hyperparameters such as the number of bees in the hive, the percentage of scouts to onlookers, and the number of top solutions to be followed. Finally, useful applications of the new algorithm are discussed, as well as limitations and possible future research opportunities.

---

<sup>1</sup>University of Kassel, Germany

<sup>2</sup>Tilburg University, Netherlands

## Corresponding Author:

Ulrich Schroeders, Psychological Assessment, Institute of Psychology, University of Kassel, Hollaendische Str. 36-38, Kassel 34127, Germany.

Email: schroeders@psychologie.uni-kassel.de

**Keywords**

bee swarm optimization, metaheuristics, structural equation modeling, dimensionality, model specification search

Metaheuristics are general-purpose algorithms that can be used to efficiently solve optimization problems that are computationally too demanding for exhaustive search solutions (Du & Swamy, 2016). Many metaheuristics are inspired by phenomena in nature, which often rely on multiple agents (i.e., population-based approaches). A prominent example is the ant colony optimization (ACO; Deneubourg et al., 1983) algorithm that mimics the foraging behavior of ants. Metaheuristics are adaptive and nondeterministic, as they converge toward optimal or near-optimal solutions by selecting and evaluating over several iterations, using information from previous iterations to adjust the selection process (Colomi et al., 1996; Dorigo & Stützle, 2010). In psychological research, metaheuristics have been used for several purposes, for example, to compile short scales that fulfill certain preset criteria (e.g., Janssen et al., 2017; Leite et al., 2008; G. A. Marcoulides & Drezner, 2001; Olaru et al., 2015, 2019; Schroeders et al., 2016a, 2016b). There is also an extensive research literature on how metaheuristics can be used to correct specification errors between a proposed model and the true model (e.g., G. A. Marcoulides & Drezner, 2003; G. A. Marcoulides et al., 1998; G. A. Marcoulides & Ing, 2012; G. A. Marcoulides & Leite, 2013; G. A. Marcoulides & Schumacker, 2001; K. M. Marcoulides & Falk, 2018). The basic idea behind this model specification search (Long, 1983) is to specify a slightly modified version of an initial, intended model and to compare both in terms of model fit (e.g., Akaike information criteria [AIC] or Bayesian information criteria [BIC]). Manually specifying such competing measurement models is feasible if the model is not too complex and the misspecification not too large, but with an increasing number of variables and possible modifications, this approach is impractical and will cover only a small to negligible part of the parameter space (K. M. Marcoulides & Falk, 2018). Instead, model specification search can be understood as a combinatorial optimization problem (G. A. Marcoulides & Drezner, 2003; G. A. Marcoulides & Schumacker, 2001). For its solution, a variety of metaheuristics (for an overview see G. A. Marcoulides & Ing, 2012) have been proposed such as ACO (Dorigo & Stützle, 2010; Leite et al., 2008), Genetic Algorithm (Holland, 1992; G. A. Marcoulides & Schumacker, 2001), Tabu Search (Drezner & Marcoulides, 1999; Glover, 1986; G. A. Marcoulides et al., 1998), and Simulated Annealing (Černý, 1985; G. A. Marcoulides & Drezner, 1999).

In the present article, we intend to complement previous efforts in model specification search using metaheuristics. Previous approaches and tools are helpful if the researcher has at least a general idea of the structure of the measurement instrument (e.g., number of underlying dimensions). However, in early stages of test or questionnaire development, little is known about the dimensional structure of a measure. Moreover, almost always item selection is necessary because more items are pilot

tested than intended for the final instrument. These two requirements—determining the factor structure and item selection—represent a considerable extension of the already challenging combinatorial problem faced in model specification searches with a known number of underlying dimensions. The traditional non-heuristic procedure in psychology is to run an exploratory factor analysis (EFA; Fabrigar & Wegener, 2011) in combination with item selection (e.g., Kruyen et al., 2013; Steger et al., 2022). This approach signifies a number of individual decisions to be made for the structure finding part such as deciding on the estimator (Preacher & MacCallum, 2003), determining the number of factors to be extracted (Preacher et al., 2013), choosing a rotation method (Schmitt & Sass, 2011), and interpreting the solution. Subsequently, items with substantial cross-loadings are excluded and these steps are repeated until a robust and interpretable solution is found. This might seem like a straightforward procedure, but usually it is not. In each iteration of the procedure, there are multiple options. For example, several methods have been suggested to determine how many factors should be retained, but without clear recommendations as to when they should be used (Auerswald & Moshagen, 2019). Thus, item selection with a simultaneous assignment of items to factors is a formidable optimization problem with many researchers' degrees of freedom (Nelson et al., 2018; Simmons et al., 2011).

To address this issue, we introduce a new population-based metaheuristic to the psychometric literature, a BSO algorithm (for an overview of previous applications in the computational sciences that used algorithms that were motivated by the behavior of bees (see Karaboga & Akay, 2009; Karaboga et al., 2012). In contrast to the foraging behavior of ants that led to the formulation of ACO algorithms, the behavioral repertoire of bees is more sophisticated due to a division of labor. *Scout bees* search for food sources in the surrounding of the hive, whereas *onlooker bees* are directed to certain locations to exploit promising food sources. The main idea of such a division of labor in model specification search is that the global search (scout bees) focuses on the identification of the number of factors and general factor-item allocation, whereas the specific search (onlooker bees) focuses on optimizing a simple structure thereof by reallocating or removing single items. We evaluate the usability and effectiveness of such a BSO in determining the factor structure and selecting items for simple structure in two real data sets representing hierarchical constructs.

## Model Specification Search in Bifactor Modeling

For the current examples, we choose a bifactor model that includes a general factor reflecting the common variance of all items and specific factors covering additional shared variance within subsets of items. The specific or nested factors are orthogonal to each other and to the general factor. Bifactor models have originally been invented in intelligence research (Holzinger & Swineford, 1937), but recently experienced a renaissance as an important structural representation of multidimensionality in various other fields (Reise, 2012; Reise et al., 2010). We chose this model because many

psychological constructs are hierarchical in nature (e.g., Kotov et al., 2021; McGrew, 2009; Moshagen et al., 2018) and bifactor models allow for a representation of this hierarchy without imposing proportionality constraints as in higher order models (Brunner et al., 2012). In comparison with a correlated factor model, a bifactor model allows items to load only on the general factor, so that the information of all items can be used, even for items that otherwise cannot be easily allocated to a specific factor.

Consider a measure with  $n$  indicators, a strong general factor, and an unknown number of specific factors. The factor model in matrix notation can be denoted as:

$$y = \Lambda\eta + \epsilon$$

where  $\eta$  denotes the latent variable,  $\Lambda$  is the factor loading matrix, and  $\epsilon$  represents the residuals. For instance, the factor model may look as follows:

$$\Lambda = \begin{bmatrix} \lambda_{11} & \lambda_{12} & 0 & \cdots & \lambda_{1n} \\ \lambda_{21} & \lambda_{22} & 0 & \cdots & \lambda_{2n} \\ \lambda_{31} & \lambda_{32} & 0 & \cdots & \lambda_{3n} \\ \lambda_{41} & 0 & 0 & \cdots & 0 \\ \lambda_{51} & 0 & \lambda_{53} & \cdots & \lambda_{5n} \\ \lambda_{61} & 0 & \lambda_{63} & \cdots & \lambda_{6n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}$$

In this example, the first three indicators load on the general factor (first column) and the first specific factor (second column). The fourth indicator only loads on the general factor, whereas the fifth and sixth indicators load on the general and the second specific factor, and so on. The last indicator is completely excluded as indicated by zero loadings in all columns. The same item-factor assignment that is given in the matrix can also be expressed as a vector of integers where each element represents an indicator and the specific value represents the assignment to a factor (for a similar notation, see G. A. Marcoulides & Drezner, 2001; Murohashi & Toyoda, 2007):

$$[1 \ 1 \ 1 \ 0 \ 2 \ 2 \ \dots \ -1]$$

Here, positive integers (e.g., 1 and 2) denote that an indicator loads on the respective specific factor (e.g., 1 and 2) and on the general factor. A value of 0 denotes an item that loads on the general factor only, and  $-1$  represents an item that is excluded from the item pool.

As BSO is a population-based metaheuristic, multiple solutions are followed simultaneously. Each possible item assignment or model is considered a location in the area and a potential food source that is visited by a bee. All models are evaluated by a predefined optimization function including multiple criteria and sorted accordingly (food quality). The better a model is evaluated (promising food source), the more bees are requested to follow and investigate alternative solutions similar to this model—in analogy to the waggle dance of honey bees. In nature, scout bees open up

new food sources, whereas onlooker bees search for food in the vicinity of previously explored, promising food sources. With respect to model specification search, scout bees initiate rather global model revisions (e.g., removing a factor), whereas onlooker bees investigate alternative models at a more fine-grained level (e.g., re-assigning a single item). In the following, we will describe the specific operations of the virtual scout and onlooker bees in more detail.

## Main Principles of Bee Swarm Optimization

We identified four key principles of bee foraging behavior and translated them to model specification search:

First, scout bees systematically explore new vicinities based on previously collected information on promising food sources that have been shared in the beehive. We allow scouts to introduce major changes to the best candidate solutions of previous iterations: (a) adding a nested factor (Scout 1 in the following example), (b) splitting a nested factor (Scout 2), (c) removing a nested factor (Scout 3), or (d) merging two nested factors (Scout 4). For instance, these operations could result in the following item assignments:

[1 1 1 0 2 2 2 2 2 0 3 3 3 3 0 4 4 4 4 4 0 4 4 4 -1] (*Best previous solution*)

[1 1 1 5 2 2 2 2 2 5 3 3 3 3 5 4 4 4 4 4 5 4 4 4 -1] (*Scout 1: Add factor*)

[1 1 1 0 2 2 2 2 2 0 3 3 3 3 0 4 4 4 4 4 0 5 5 5 -1] (*Scout 2: Split factor*)

[1 1 1 0 2 2 2 2 2 0 0 0 0 0 0 4 4 4 4 4 0 4 4 4 -1] (*Scout 3: Remove factor*)

[1 1 1 0 2 2 2 2 2 0 2 2 2 2 0 4 4 4 4 4 0 4 4 4 -1] (*Scout 4: Merge factor*)

Importantly, each scout *randomly* decides which of the four operations is applied to the best previous solution. If there are multiple possible ways to execute an operation (e.g., which factor could be split), this decision is also made randomly.

Second, onlookers search for food in the immediate proximity of a good food source. In psychometric terms, onlooker bees introduce minor changes to the model that may involve (a) adding an item to a nested factor (if it was not assigned to a nested factor previously, Onlooker 1), (b) removing an item from a nested factor (so that it only loads on the general factor, Onlooker 2), (c) swapping an item between nested factors (Onlooker 3), or (d) completely removing an item from the pool (Onlooker 4). Again, the decision of which these operations are executed and which item is affected is made randomly. For example, minor tweaks may look like these item-factor changes:

[1 1 1 0 2 2 2 2 2 0 3 3 3 3 0 4 4 4 4 4 0 4 4 4 -1] (*Best previous solution*)

[1 1 1 1 2 2 2 2 2 0 3 3 3 3 0 4 4 4 4 4 0 4 4 4 -1] (*Onlooker 1: Adding item*)

[1 1 1 0 2 2 2 2 2 0 3 3 3 3 0 4 4 4 4 4 0 4 4 4 -1] (*Onlooker 2: Removing item*)

[1 1 1 0 2 2 3 2 2 0 3 3 2 3 0 4 4 4 4 4 0 4 4 4 -1] (*Onlooker 3: Swapping items*)

[1 1 1 0 2 2 2 2 2 2 0 3 3 3 3 0 4 4 4 4 4 0 4 4 -1 -1] (*Onlooker 4: Dropping item*)

Third, successful scouts provide information about the location of good food sources (e.g., direction, distance, and quality). The more promising the source, the more onlooker bees are attracted. In psychometric terms, all measurement models are evaluated according to a user-defined optimization function and recorded in a sorted list. In analogy to more profitable food sources leading to intensive dances attracting more onlookers, models that better fulfill the predefined criteria will entail more intense local searches (e.g., 10 onlookers for the best model and eight onlookers for the second best model). Moreover, to streamline the specification search we also included a Tabu list ensuring that the same model is not estimated several times (which may be conceived as depletion of the food source).

Fourth, after each iteration, the list of the best models is updated and used as a starting point by both scout and onlooker bees in the next iteration. To reduce the possibility of focusing too much on only one specific region in the search space (i.e., reaching a local optimum), models that are on top of the best model list are no longer pursued after a predefined number of iterations, which is equivalent to the depletion of food sources. Please note that these models are not removed from the list; they are simply not further explored. Pseudocode for the complete BSO procedure can be found in the online supplement.

## **Advantages of Bee Swarm Optimization**

BSO has several advantages in test or questionnaire development compared with other methods (see Table 1): First, the algorithm can perform item selection during the optimization. Especially in the early stages of test development, the question of dimensionality underlying a questionnaire is often complicated by the need to select appropriate items from a larger pool. Both exploratory and confirmatory bifactor analyses are not suitable to conduct item selection out-of-the-box. Often items are selected after the structure of the questionnaire was established, for example, by removing items that contributed most to model misfit. However, the dimensionality depends on the specific abbreviated item set, which is why both analysis steps should be conducted simultaneously. Greedy stepwise selection mechanisms also depend on the sequence of item removal; thus, they are prone to run into local optima. In contrast, BSO evaluates item sets during the optimization in a non-greedy manner, that is, it can recover items that would be excluded if only the best solution would have been followed.

Second, although BSO is logically an exploratory, data-driven approach (Wagenmakers et al., 2012), its models are evaluated using confirmatory factor analyses, allowing to find an unambiguous structure for a measure. The various structure-finding or structure-testing procedures—exploratory factor analysis with bifactor rotation (bifactor EFA), BSO, bifactor exploratory structure equation modeling (BESEM) with target rotation, and confirmatory factor analysis with bifactor rotation (bifactor CFA)—can be ordered on a continuum from exploratory to confirmatory, from no to clear specifications regarding the structure of a measure. Bifactor

**Table 1.** Comparison of BSO Algorithm With Exploratory and Confirmatory Bifactor Modeling.

BSO	Exploratory bifactor	Confirmatory bifactor
Item selection possible	No item selection included	No item selection included
Data-driven	Data-driven	Theory-driven
Simple structure	Cross-loadings Biased estimates	Strong theory needed Anomalous results
No sequence effect		+ Item selection
Multiple criteria		Sequence effects One criterion at a time

*Note.* BSO = Bee swarm optimization. Exploratory/confirmatory bifactor = EFA/CFA with bifactor rotation.

EFA and BSO are clearly exploratory, whereas bifactor CFA is completely confirmatory. BESEM with target rotation is semi-confirmatory (see Huang et al., 2017), because it combines features of EFA and CFA: On one hand, they allow the occurrence of cross-loadings, and on the other hand, they are also based on a prespecified model structure (Marsh et al., 2010; Morin et al., 2016). BESEM might be useful to take into account the fuzziness of indicators in many real-world settings, but for the target rotation, it is necessary to specify a target structure in advance. More knowledge about a measure will often lead to better models, but in early stages of test development such knowledge is lacking. This is when BSO can prove useful, because the algorithm makes no prior assumptions about the assignment of items to factors and about the number of factors. Nonetheless, BSO enforces a simple structure by only allowing one additional factor loading besides the loading on the general factor, restricting all other loading to zero. This leads to more purified specific factors that might be easier to interpret. However, this clear assignment of items to factors by BSO is not achieved by requiring strong theoretical assumptions as in the case of confirmatory factor analyses.

Third, with respect to item selection, traditional methods rely on stepwise exclusion of items according to a single criterion, for example, deleting the item with the lowest item-total correlation (Krueger et al., 2013) or the one that contributes most to model misspecification (Whittaker, 2012). Usually, this item deletion is based on a single metric (Clifton, 2020; Steger et al., 2022). In contrast, model search of BSO can incorporate several criteria simultaneously (e.g., model fit, reliability, and length of the scale). The simultaneous inclusion of different optimization goals is a major advantage of metaheuristics in general (Dorigo & Stützle, 2010; Olaru et al., 2015, 2019). For example, metaheuristics have been used to create reliable and cross-culturally invariant short versions of a Big Five personality inventory (Jankowsky et al., 2020). The consideration of different, possibly conflicting criteria in BSO prevents a one-sided optimization at the expense of other criteria (e.g., the attenuation paradox by Loevinger, 1954, and Schroeders et al., 2016b).

Fourth, bifactor models might run into estimation problems or produce anomalous results. In case of exploratory bifactor modeling, Schmid–Leiman and alternative bifactor rotations lead to inaccurate parameter estimates when the specified model departs from perfect cluster structure (e.g., item cross-loadings on nested factors; Mansolf & Reise, 2016). In case of confirmatory bifactor modeling, overfactoring might lead to anomalous results. Therefore, Eid et al. (2017) proposed two theory-driven alternative models, the bifactor-(S–1) model and the bifactor-(S.I–1) model, in which either a factor or a single item per factor is set as a reference. In case of clear theoretical assumptions, such reference models are the preferred choice, but without a clear understanding of the factor structure, these models are difficult to specify. In contrast to these bifactor models with a reference, BSO makes no prior assumption how items should be assigned to factors, which is completely subjected to the optimization process. Thus, BSO can circumvent the issue of overfactoring by allowing items to load on the general factor only.

Fifth, in comparison with an exhaustive search, BSO is efficient. The combination of structure finding and item selection is a huge combinatorial problem (G. A. Marcoulides & Drezner, 2001). To give an idea on the combinatorial complexity, let us assume a bifactor model with 25 items, and, for the sake of simplicity, we restrict the number of nested factors to a maximum of two. Each item can thus take one of four states (item is excluded, loads on the general factor only, loads on either the first or second nested factor), which results in a total of approximately  $4^{25} = 1,125,899,906,842,624$  models (if models with less than three items per factor are counted). If we estimate that each model takes one-tenth of a second to compute, then even with parallel computation on 10 cores, the analysis of all models would take 356,776 years. In comparison, BSO achieves results in considerably less time in similar settings (as in the following empirical examples)—usually between 1 and 3 hr.

## Research Objectives

The aims of the present proof-of-principle study are twofold: First, we will systematically study the influence of different hyperparameters such as the number of scout bees and onlooker bees, respectively, on the convergence and the quality of the provided solutions. For this purpose, we used the well-known Holzinger–Swineford data set to specify a bifactor model with all subtests loading on a general factor and in addition on two to five domain-specific, nested factors. The task of the BSO algorithm is to group the indicators into nested factors where the number of nested factors is not specified further. The optimization follows a predefined function that takes into account different criteria such as model fit, minimal factor loading, and the total number of items. The results will be compared with the results of exploratory and confirmatory factor analyses with Schmid–Leiman rotation and a varying number of nested factors.



Second, we study the usability and applicability of BSO in another data set featuring a (SD3, Jones & Paulhus, 2014). In the last decade, personality traits that were associated with ethically, morally, and socially questionable behavior have received a lot of attention among personality researchers (e.g., Moshagen et al., 2018). The SD3 consists of 27 items with nine items each for machiavellianism (a manipulative attitude), narcissism (excessive self-love), and psychopathy (lack of empathy). Because the factor structure of the SD3 has not been satisfactorily clarified (Jones & Paulhus, 2014; Persson et al., 2019), we use the questionnaire as an example to showcase the possibilities of the BSO algorithm. More precisely, we demonstrate how to derive essentially equivalent models that all satisfy the predefined psychometric criteria, thus, both streamlining the item pool of a measure and exploring its structure.

## Method

### Data sets

*Holzinger–Swineford Data Set.* We used the Holzinger–Swineford data set, which is a classic textbook example that has been reanalyzed many times in the psychometric literature (e.g., Murohashi & Toyoda, 2007). The data set contains intelligence test scores for 24 cognitive tests from 301 children attending two schools (for more information, see Holzinger & Swineford, 1939). The study on which the data are based has become famous in the assessment literature because it was the first application of a bifactor model (Holzinger & Swineford, 1937).

Concerning the structure of the Holzinger–Swineford data set, the 24 subtests were chosen from five different broad abilities (number of subtests in parentheses): (a) spatial abilities (4), (b) verbal abilities (5), (c) mental speed (4), (d) memory (6), and (e) mathematical deduction (5). Instead of the intended factor structure, Holzinger and Swineford (1939) concluded that a bifactor model with four specific factors (without the last factor) better fitted the data.

*Dark Triad Data Set.* One of the most influential questionnaires to measure the dark core of personality is the SD3, in which previously separate developmental strands are merged by combining and abbreviating existing measures (Jones & Paulhus, 2014). Data were collected online and distributed under an Open Database License.<sup>1</sup> The data set consists of the responses of 18,192 participants on the 27 items of the SD3.

With respect to the structure of the measure, the questionnaire developers described the fit of the three-dimensional model in the original publication as “not impressive (RMSEA [Root Mean Square Error of Approximation] = .07, CFI [Comparative Fit Index] = .82, and TLI [Tucker-Lewis-Index] = .80)” (Jones & Paulhus, 2014, p. 33). Consequently, Persson et al. (2019) investigated the psychometric properties of the SD3 in three independent samples and proposed a bifactor structure with an overarching dark personality as general trait and either three nested factors or two nested factors, in which machiavellianism and psychopathy collapse, while narcissism is modeled as a second independent factor.

**Table 2.** Hyperparameters of the BSO Algorithm.

Argument	Explanation	Example
n_bees	Number of bees in the hive	200
percent_scouts	Proportion of scout bees	.25
top_best	Number of top solutions to be examined	20
max_iter	Maximum number of iterations without improvement	10
min_nest_fac	Minimum number of (nested) factors	1
max_nest_fac	Maximum number of (nested) factors	5
depletion	Number of iterations after which the food source is depleted	5

Note. Arguments that are varied in the empirical illustration are given in the upper part of the table.

### *Hyperparameters of the Bee Swarm Optimization Algorithm*

Metaheuristics often provide efficient solutions to a problem, but not necessarily the best solution. The convergence behavior of the BSO depends on the optimization function and several hyperparameters regulating the search behavior and restricting the search space (the most important are given in Table 2). To study their influence in more detail, we systematically varied the following three hyperparameters in the Holzinger–Swineford data set: (a) the number of bees in the hive (n\_bees), (b) the percentage of scouts among all bees (percent\_scouts), and (c) the number of top solutions that attract onlooker bees (top\_best). n\_bees should have a direct influence on the convergence behavior and on the duration of the search: The more bees, the larger the area that can be covered and the more extensive the search. The other two hyperparameters (percent\_scouts and top\_best) affect the relation of diversification to intensification. The higher the proportion of scouts among all bees, the more diversification occurs compared with intensification. The top\_best parameter determines how many promising solutions are intensified. A small number indicate a strong intensification but also a more specific local search.

We chose the Holzinger–Swineford data set for this demonstration because researchers have a good understanding of the structure (e.g., Murohashi & Toyoda, 2007). In more detail, we used the following design for the empirical illustration of the hyperparameters: n\_bees with six levels (100, 200, 400, 600, 800, and 1000), percent\_scouts with three levels (.25, .50, and .75), and top\_best with three levels (10%, 20%, and 50% of n\_bees). Each combination was run across 20 seeds, totaling  $6 \cdot 3 \cdot 3 \cdot 20 = 1080$  runs. Because the overall best solution is essentially unknown, we considered the overall best solution across all runs according to our optimization criteria the best solution.

### *Statistical Analysis*

The BSO algorithm is openly available from Github: <https://github.com/FlorianScharf/BSO>. BSO relies mainly on the parallel package for parallel

computing (v.4.2.2; R Core Team, 2022), lavaan for CFA (v.0.6.13; Rosseel, 2012), and ggplot2 for plotting the convergence plots (v.3.4.0; Wickham, 2016). All syntax and data files necessary to reproduce the analyses of this study can be found in an online repository: <https://osf.io/kx4fg/>.

### The Optimization Function

The core of a metaheuristic search is the optimization function that evaluates the quality of the models. The optimization function is modular in design and often contains several criteria. In the present context, we focused on model fit, factor loadings, and the number of items included in the final model. All criteria were logit-transformed to scale the outcomes of the different formulae (e.g., Equations 1 and 2) to the range between 0 and 1 to maximally differentiate between models around a preset cutoff value (Janssen et al., 2017; Schroeders et al., 2016a). We labeled these value  $\nu$  in reference to the Greek word nectar ( $\nu\epsilon\kappa\tau\alpha\rho$ ).

With respect to the first criterion, model fit, we used a combination of an incremental fit index, the Comparative Fit Index ( $CFI \geq .95$ ), and an absolute fit index, the Root Mean Square Error of Approximation ( $RMSEA \leq .06$ )—as proposed with the two-index presentation strategy (Hu & Bentler, 1999). All models had continuous indicators, which is why the maximum likelihood estimator was used. To streamline the convergence of the BSO algorithms by reducing the number of non-promising models, both observed and latent variances were forced to be non-negative (bounds = “pos.var”). The chosen cutoff values for the fit indices constitute the inflection point of the function.

$$\nu_{CFI} = \frac{1}{1 + \exp(50 \cdot (.95 - CFI))} \quad (1)$$

$$\nu_{RMSEA} = \frac{1}{1 + \exp(-50 \cdot (.06 - RMSEA))} \quad (2)$$

The second criterion of the optimization function deals with the minimal absolute factor loading across all nested factors, which was set to  $\lambda = .30$ .

$$\nu_{\lambda} = \frac{1}{1 + \exp(15 \cdot (.30 - |\lambda_{min}|))} \quad (3)$$

The third criterion deals with the number of indicators that will be used in the final version. The indicators in both data sets had been extensively tested; that is, item selection had already taken place in previous stages of test development. For example, the original version of SD3 consisted of 41 items from which the final 27 items were drawn. For this reason, we choose a cutoff of 95% of all items, which essentially prevents any substantial item selection.

$$\nu_{items} = \frac{1}{1 + \exp(30 \cdot (.95 - n(items)))} \quad (4)$$

Please note that the chosen cutoffs slightly differ in the second example to accommodate for the different data set, that is, for the dark triad data set we deemed a CFI  $\geq .90$  satisfactory and strived for a minimal absolute factor loading  $\geq .20$ . The overall weighting of the three criteria in the optimization function favors the model fit:

$$\nu_{total} = 2 \cdot (\nu_{CFI} + \nu_{RMSEA}) + \nu_{\lambda} + 2 \cdot \nu_{items} \quad (5)$$

Models in which one of the criteria was below a predefined threshold of  $\nu \leq .0001$  were not followed up. Please note that other criteria could also easily be implemented in the optimization function such as maximizing correlations with covariates, favoring item sets that offer measurement invariance with respect to grouping variables, balancing content domains, etc. In this introduction to BSO, we will adhere to some basic principles of test construction.

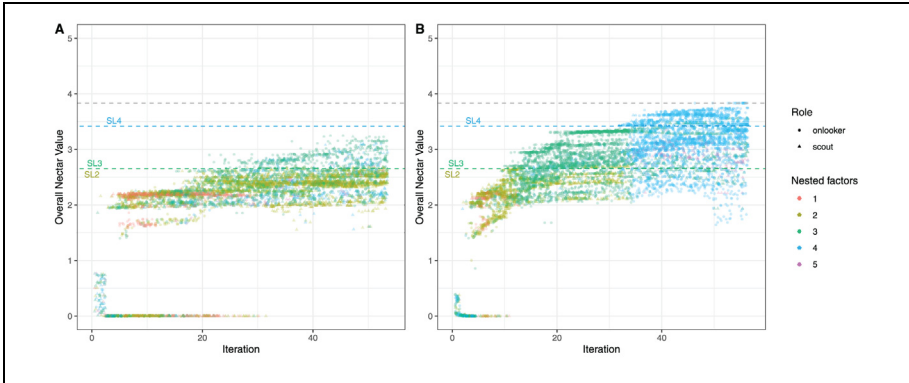
## Results

### Example 1: Holzinger–Swineford Data Set

Specifying the originally intended bifactor model with five nested factors (Holzinger & Swineford, 1939) resulted in a confirmatory factor model with acceptable fit:  $\chi^2(228, N = 301) = 438.9, p < .001, CFI = .922, RMSEA = .055$ . However, the standardized loadings of the fifth factor (deduction) fluctuated unsystematically around 0, indicating overfactoring. A confirmatory bifactor model with four specific factors and the indicators of deduction set as reference overcame these problems:  $\chi^2(233, N = 301) = 444.0, p < .001, CFI = .922, RMSEA = .055$ . With respect to the optimization function which was the base for the BSO algorithm, this model yielded a  $\nu$ -value of 3.19.

As another point of comparison, we conducted exploratory factor analyses and Schmid–Leiman rotation with two, three, and four specific factors (i.e., SL2, SL3, and SL4). In a subsequent step, we assigned all items to the general factor only that had a factor loading below .30 on the nested factor to make a reasonable comparison with the BSO. The bifactor solution with two nested versus three nested factors came to similar values in the optimization function ( $\nu_2 = 2.65$  vs.  $\nu_3 = 2.65$ ), whereas the solution with four nested factors scored higher ( $\nu_4 = 3.42$ ).

The best model derived by the BSO algorithm was a bifactor model and four specific factors, resulting in a substantially higher value in the optimization function ( $\nu = 3.88$ ). Compared with the intended structure, there were some modifications: One indicator of the memory factor (*figurew*) only loaded on the general factor and two indicators that were originally intended for the last factor (*numeric* and *arithmet*) loaded on the mental speed factor instead. This last modification seems reasonable because speed played a crucial role in solving these simple arithmetic problems. The model fit was slightly better:  $\chi^2(232, N = 301) = 414.6, p < .001, CFI = .933,$



**Figure 1.** Convergence Plots of Two Exemplary BSO Runs.

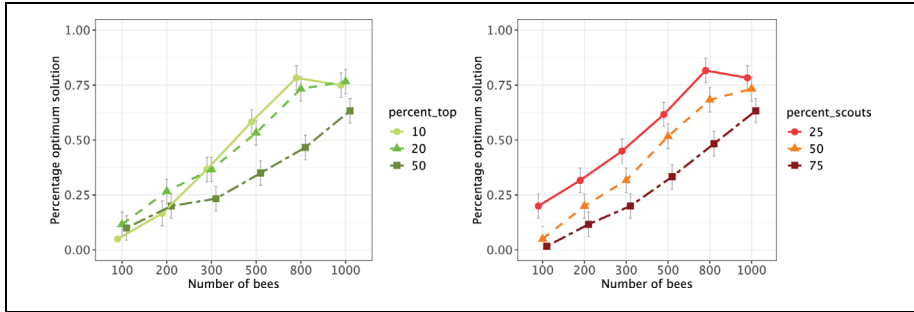
*Note.* In both runs, the same optimization function was used, but with different hyperparameters. Left panel:  $n\_bees = 120$ ,  $top\_best = 60$ ,  $percent\_scouts = .75$ . Right panel:  $n\_bees = 120$ ,  $top\_best = 24$ , and  $percent\_scouts = .25$ . Colored dashed lines marked with SL2, SL3, and SL4 represent the nectar values of the confirmatory bifactor model with two, three, and four specific factors and Schmid–Leiman (SL) rotation. Nectar values of SL2 and SL3 are almost identical. The gray dashed line indicates the empirical optimum. The syntax can be found online at <https://osf.io/jvrpq>.

RMSEA = .051 (parameter estimates for the four-dimensional solutions are given in the online supplement). Given the constraints specified in the optimization function, this solution was deemed the optimal solution in the context of the subsequent illustration.

Figure 1 shows the convergence plots of two exemplary specifications of the BSO algorithm with the same optimization function and a total of 120 bees. In the left panel, the top 60 solutions were pursued in each iteration, whereas only the best 24 solutions were used in the right panel. In addition, the percentage of scouts differed across the two examples: 50% versus 25%. As a consequence, the process in the left panel was less diversified and ran into a local minimum, while the optimum solution was found in the right panel. This led to the question which hyperparameter settings have a positive influence on the convergence behavior.

### *Systematic Variation of Hyperparameters*

Overall, the algorithm provided consistent and efficient results in the Holzinger–Swineford data set, because even with a small number of bees the optimal solution was often found across 20 seeds. With 800 or more bees, the optimal solution was provided in almost all cases (for more detailed results, see online supplement). To determine the influence of the hyperparameters on the consistency of the results, we conducted a three-way analysis of variance (ANOVA) with the hyperparameters as independent variables and the frequency of how often the best solution was found across 20 seeds as the dependent variable. The main effects of all three hyperparameters were significant, with the



**Figure 2.** Proportion of Optimal Solutions Among Random Starts as a Function of Hyperparameters.

Note. Points represent the average proportion of optimal solutions across all random seeds per condition, and error bars represent  $\pm 1$  standard error.











number of bees being the most influential factor,  $F(5, 1026) = 63.99, p < .001, \eta_p^2 = .24$ . Figure 2 indicated improved consistency of results if only 10% or 20% of the best models were followed up in subsequent iterations,  $F(2, 1026) = 10.96, p < .001, \eta_p^2 = .02$ . Moreover, using only 25% scouts (and 75% onlookers) yielded better results than the opposite ratio,  $F(2, 1026) = 27.78, p < .001, \eta_p^2 = .05$ . In summary, at least for this example, it is important for the BSO to weigh intensification more strongly than diversification. Also, the interaction between the percentage of scouts and the number of top solutions to be followed was significant (for the complete results of the ANOVA, please see online supplement).

### Example 2: Dark Triad Data Set

The intended confirmatory bifactor model with nine items loading on each of the three specific factors machiavellianism, psychopathy, and narcissism (Jones & Paulhus, 2014) had sufficient model fit,  $\chi^2(297, N = 18,192) = 17,175.1, p < .001$ , CFI = .901, RMSEA = .057. However, some of the factor loadings on the nested factors turned out to be low, that is, three were below .10 and eight below .20. The factor saturation of the nested factors was low for machiavellianism ( $\omega_M = .11$ ) and for psychopathy ( $\omega_M = .19$ ), and substantial for narcissism ( $\omega_M = .44$ ). The  $\nu$ -value of this theoretically established model was  $\nu = 3.87$ .

Table 3 gives the best solutions of 10 independent seeds, out of which seven yielded higher  $\nu$ -values than the intended model. The best model had essentially the same model fit as the theoretical model, but a slightly higher minimal factor loading on the nested factors ( $\lambda_{min} = .108$  vs.  $\lambda_{min} = .074$ ; for all factor loadings of the best BSO model, see online supplement). The right side of the table shows the color-coded item-factor assignment. Overall, the structure found by the BSO algorithm matches the intended structure. There were two deviations of the best solution from

**Table 3.** The Best BSO Solutions of 10 Seeds Concerning the Structure of the SD3.

Seed	$\nu$	CFI	RMSEA	$\lambda_{min}$	Item-factor assignment
9	3.968	.902	.056	.108	
1	3.935	.902	.056	.092	
4	3.935	.902	.056	.092	
5	3.935	.902	.056	.092	
7	3.935	.902	.056	.092	
8	3.935	.902	.056	.092	
2	3.907	.902	.056	.078	
6	3.698	.893	.058	.106	
10	3.637	.890	.059	.115	
3	3.633	.892	.059	.097	

M1 M2 M3 M4 M5 M6 M7 M8 M9 N1 N2 N3 N4 N5 N6 N7 N8 N9 P1 P2 P3 P4 P5 P6 P7 P8 P9

Note. In all runs, the same optimization function was used with  $CFI \geq .90$ ,  $RMSEA \leq .06$ ,  $|\lambda_{min}| \geq .20$ , and item selection essentially turned off. Hyperparameters: n\_bees = 2000, top\_best = 200, percent\_scouts = .25. M1-M9 = machiavellianism; N1-N9 = narcissism; P1-P9 = psychopathy; red/orange/yellow = nested factors; light gray = general factor only.

the intended structure: First, two machiavellianism items loaded on the general factor only (see light gray squares). Second, the item “*Whatever it takes, you must get the important people on your side*” was assigned to narcissism instead of machiavellianism (with the item arguably measuring grandiosity). Both modifications can be understood as a sign that the specificity of the machiavellianism items was lower.

### Discussion

In this article, we introduced a new metaheuristic—BSO—to the psychometric literature. In the course of measurement development, behavioral researchers often have to make several decisions, such as determining the number of factors to extract, selecting items, and thinking about alternative assignments of items to factors—rendering the process a formidable optimization problem. BSO tackles the question of dimensionality of a measurement instrument in combination with a possible item selection. Both structure finding and the item selection are conducted jointly according to an optimization function that may include several criteria (e.g., model fit and factor saturation). In contrast to exploratory bifactor analysis, BSO avoids biased estimates (Mansolf & Reise, 2017) and renounces cross-loadings that sometimes complicate the interpretation of factors. In contrast to confirmatory bifactor analysis, BSO does not presuppose strong theoretical assumptions and avoids the sometimes occurring anomalous results (Eid et al., 2017). With regard to the extensive metaheuristic literature concerning model specification search (G. A. Marcoulides & Drezner, 2003; G. A. Marcoulides et al., 1998; G. A. Marcoulides & Ing, 2012; G. A.

Marcoulides & Leite, 2013; G. A. Marcoulides & Schumacker, 2001; K. M. Marcoulides & Falk, 2018), both similarities and differences to BSO can be found. One obvious commonality is that metaheuristics are often based on biological mechanisms of an animal population (e.g., a fish swarm or an insect colony). In addition, many metaheuristics try to strike a balance between two opposing strategies—diversification (or exploration) versus intensification (or exploitation)—and they will be more successful in finding a near-to-optimal solution to a given optimization problem if a (dynamic) balance between both strategies can be established (Blum & Roli, 2003). For the model specification search discussed in the current article, measurement and structural parameters of a structure equation modeling (e.g., factor loadings and factor correlations) are conceptualized as a vector that is subject to a random modification (see G. A. Marcoulides & Drezner, 2003; G. A. Marcoulides & Schumacker, 2001). The resulting model is then evaluated regarding predefined criteria which in subsequent iterations influences the probability of items to be assigned to factors. Even though the algorithms are similar on this abstract level and often achieve similar results, there are still major differences between the various algorithms (G. A. Marcoulides & Ing, 2012; Olaru et al., 2015; Schroeders et al., 2016b). If one wants to draw parallels between ACO and BSO, one could link the local search of the onlooker bees to the mode of operation of the ants, whereas the global search of the scout bees has no direct equivalent. In ACO, exploration and exploitation depend on the number of ants, search duration, and evaporation of pheromones. In BSO, however, the two processes are split with scouts engaging in diversification and onlookers in intensification. Such a division of labor might favor a dynamic balance between both strategies when it comes to finding a measure's structure. Compared with metaheuristics that have previously been used, BSO brings two major features: First, the number of factors does not have to be specified a priori, and second, the search for a solution can be combined with item selection. Thus, we deem BSO a flexible and promising new tool in the method toolbox of psychometricians.

The measurement model that is finally presented in an article is often equivalent in terms of its psychometric properties (e.g., model fit and factor saturation) to other models. Such alternative models might offer substantively meaningful concurrent explanations to the data (MacCallum et al., 1993). In practice, however, the number of models to be manually specified and tested is reduced to a small subset of models that are theoretically motivated. If a researcher tries to specify all conceivable models, an iterative, knowledge-driven search is neither practical nor exhaustive (G. A. Marcoulides & Ing, 2012; K. M. Marcoulides & Falk, 2018). BSO offers a different solution to this problem: BSO systematically searches the solution space of competing measurement models, so that researchers get a sense of the variation of possible good (not essentially equivalent) solutions. In this sense, BSO is a good complement to more theory-driven methods to assess the robustness of effects such as specification curve analysis (Simonsohn et al., 2020) or multiverse analysis (Steege et al., 2016). In contrast, the BSO algorithm provides the researcher with a set of appropriate, but data-driven models that need to be cross-validated before claiming real



validity (G. A. Marcoulides & Leite, 2013). Based on this set, alternative models can be evaluated with respect to their usefulness and theoretical plausibility. As exemplified with the SD3 data set, the BSO can thus point to weaknesses in test development (i.e., low specificity of the machiavellianism items).

The present article is a first-time proof of principle, which implies several limitations but also possible future research avenues. First, our presentation of BSO is currently limited to a bifactor model; however, it should be easy to extend the algorithm to other models such as a correlated factor model or a higher order model. The general idea of scout bees which introduce major changes to the model specification search and onlooker bees which engage in minor tweaks could be useful for several models. Second, we deliberately decided to leave the addition/removal of factors (or items to factors) to chance. From the convergence behavior in the examples, we conclude that the algorithm efficiently leads to a stable final solution within a few dozen iterations. Conversely, the specification search could also be more restrictive or goal-oriented, that is, the algorithm presumably can be further optimized at the expense of losing the connection to the biological model. For example, items could be added/removed with Tabu search (for a successful combination of ACO and Tabu search, see Jing et al., 2022) or the items subject to change could be selected based on their contribution to the model misfit (e.g., modification indices). Third, we tried to closely mirror the foraging behavior of honey bees, but also made simplifications to the rather complex communication of bees. Thus, one might question specific decisions we made in translating the biological principles into psychometric concepts. We acknowledge that the chosen approach is not the only conceivable implementation of a BSO algorithm. But just as there are different types of bees (with a diverse behavioral repertoire), there are also different BSO algorithms that could be developed (for a critical discussion of the use of metaphors in the development of new methods, we recommend Sörensen, 2015). The general idea of scout bees that introduce major changes to the model specification search and onlooker bees that engage in minor tweaks could also be useful for a wide range of applications. At this point of algorithmic development, we also know little about whether the presented set of results on hyperparameters generalizes to other data sets. Presumably, the degree to which intensification or diversification leads to better performance of BSO depends on the nature of the solution space and the optimization function. We found that 25% scouts led to better results than a higher percentage of scouts, which is interestingly backed up by observations of honeybees for which the number of scouts varies between about 5% and 35%, depending on forage availability (Seeley, 1983). To answer such questions concerning the hyperparameter more systematically, it is necessary to conduct a simulation study with different optimization problems, which is beyond the scope of this introduction.

On a more general methodological stance, we think we have just tapped the potential of metaheuristics in psychometrics. Similar optimization problems are widespread: For example, constructing parallel test forms or booklets with many boundary conditions (e.g., average item difficulty, comparable test information curves, and

predefined item coverage) is a common and reoccurring issue in educational large-scale assessment and represents another optimization problem that could also be handled with metaheuristics (see van der Linden, 2005). In addition, current applications of metaheuristics often evolve around the items of a measure (e.g., constructing short scales), but it is also possible to apply the same logic to the person side, that is, metaheuristics could be used to form groups of persons (rather than item bundles). In short, potential new applications in psychology are manifold, as well as the possibilities for further methodological developments. Therefore, we welcome the recent efforts and advances of other researchers to capitalize on the possibilities of metaheuristics such as combining ACO and Tabu search for model specification (Jing et al., 2022) or using ACO to automatically search for sensitivity parameters (Leite et al., 2022).

### **Acknowledgments**

We thank Kristin Jankowsky, Johannes Zimmermann, and our journal club for providing valuable feedback on previous drafts of this article.

### **Declaration of Conflicting Interests**

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### **Funding**

The authors received no financial support for the research, authorship, and/or publication of this article.

### **Ethical Approval**

We confirm that the work adheres to the American Psychological Association's Ethical Principles of Psychologists and Code of Conduct.

### **Supplemental Material**

Supplemental material for this article is available online.

### **Note**

1. <https://www.kaggle.com/datasets/lucasgreenwell/short-dark-triad-responses>

### **References**

- Auerswald, M., & Moshagen, M. (2019). How to determine the number of factors to retain in exploratory factor analysis: A comparison of extraction methods under realistic conditions. *Psychological Methods, 24*(4), 468–491. <https://doi.org/10.1037/met0000200>
- Blum, C., & Roli, A. (2003). Metaheuristics in combinatorial optimization: Overview and conceptual comparison. *ACM Computing Surveys (CSUR), 35*(3), 268–308. <https://doi.org/10.1145/937503.937505>

- Brunner, M., Nagy, G., & Wilhelm, O. (2012). A tutorial on hierarchically structured constructs. *Journal of Personality, 80*(4), 796–846. <https://doi.org/10.1111/j.1467-6494.2011.00749.x>
- Černý, V. (1985). Thermodynamical approach to the traveling salesman problem: An efficient simulation algorithm. *Journal of Optimization Theory and Applications, 45*(1), 41–51. <https://doi.org/10.1007/BF00940812>
- Clifton, J. D. W. (2020). Managing validity versus reliability trade-offs in scale-building decisions. *Psychological Methods, 25*(3), 259–270. <https://doi.org/10.1037/met0000236>
- Colomi, A., Dorigo, M., Maffioli, F., Maniezzo, V., Righini, G., & Trubian, M. (1996). Heuristics from nature for hard combinatorial optimization problems. *International Transactions in Operational Research, 3*(1), 1–21. <https://doi.org/10.1111/j.1475-3995.1996.tb00032.x>
- Deneubourg, J. L., Pasteels, J. M., & Verhaeghe, J. C. (1983). Probabilistic behaviour in ants: A strategy of errors? *Journal of Theoretical Biology, 105*(2), 259–271. [https://doi.org/10.1016/S0022-5193\(83\)80007-1](https://doi.org/10.1016/S0022-5193(83)80007-1)
- Dorigo, M., & Stützle, T. (2010). Ant colony optimization: Overview and recent advances. In M. Gendreau & J.-Y. Potvin (Eds.), *Handbook of metaheuristics* (pp. 227–263). Springer US.
- Drezner, Z., & Marcoulides, G. A. (1999). Tabu search model selection in multiple regression analysis. *Communications in Statistics—Simulation and Computation, 28*(2), 349–367. <https://doi.org/10.1080/03610919908813553>
- Du, K.-L., & Swamy, M. N. S. (2016). *Search and optimization by metaheuristics*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-41192-7>
- Eid, M., Geiser, C., Koch, T., & Heene, M. (2017). Anomalous results in G-factor models: Explanations and alternatives. *Psychological Methods, 22*(3), 541–562. <https://doi.org/10.1037/met0000083>
- Fabrigar, L. R., & Wegener, D. T. (2011). *Exploratory factor analysis*. Oxford University Press.
- Glover, F. (1986). Future paths for integer programming and links to artificial intelligence. *Computers & Operations Research, 13*(5), 533–549. [https://doi.org/10.1016/0305-0548\(86\)90048-1](https://doi.org/10.1016/0305-0548(86)90048-1)
- Holland, J. H. (1992). *Adaptation in natural and artificial systems* (Reprint ed.). A Bradford Book.
- Holzinger, K. J., & Swineford, F. (1937). The Bi-factor method. *Psychometrika, 2*(1), 41–54. <https://doi.org/10.1007/BF02287965>
- Holzinger, K. J., & Swineford, F. (1939). *A study in factor analysis: The stability of a bifactor solution* (Supplementary Educational Monograph, no. 48). University of Chicago Press.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Huang, P.-H., Chen, H., & Weng, L.-J. (2017). A penalized likelihood method for structural equation modeling. *Psychometrika, 82*(2), 329–354. <https://doi.org/10.1007/s11336-017-9566-9>
- Jankowsky, K., Olaru, G., & Schroeders, U. (2020). Compiling measurement invariant short scales in cross-cultural personality assessment using ant colony optimization. *European Journal of Personality, 34*(3), 470–485. <https://doi.org/10.1002/per.2260>
- Janssen, A. B., Schultze, M., & Grötsch, A. (2017). Following the ants: Development of short scales for proactive personality and supervisor support by ant colony optimization. *European Journal of Psychological Assessment, 33*(6), 409–421. <https://doi.org/10.1027/1015-5759/a000299>

- Jing, Z., Kuang, H., Leite, W. L., Marcoulides, K. M., & Fisk, C. L. (2022). Model specification searches in structural equation modeling with a hybrid Ant Colony Optimization algorithm. *Structural Equation Modeling: A Multidisciplinary Journal*, 29(5), 655–666. <https://doi.org/10.1080/10705511.2021.2020119>
- Jones, D. N., & Paulhus, D. L. (2014). Introducing the Short Dark Triad (SD3): A brief measure of dark personality traits. *Assessment*, 21(1), 28–41. <https://doi.org/10.1177/1073191113514105>
- Karaboga, D., & Akay, B. (2009). A survey: Algorithms simulating bee swarm intelligence. *Artificial Intelligence Review*, 31(1–4), 61–85. <https://doi.org/10.1007/s10462-009-9127-4>
- Karaboga, D., Gorkemli, B., Ozturk, C., & Karaboga, N. (2012). A comprehensive survey: Artificial bee colony (ABC) algorithm and applications. *Artificial Intelligence Review*, 42(1), 21–57. <https://doi.org/10.1007/s10462-012-9328-0>
- Kotov, R., Krueger, R. F., Watson, D., Cicero, D. C., Conway, C. C., DeYoung, C. G., Eaton, N. R., Forbes, M. K., Hallquist, M. N., Latzman, R. D., Mullins-Sweatt, S. N., Ruggero, C. J., Simms, L. J., Waldman, I. D., Waszczuk, M. A., & Wright, A. G. C. (2021). The Hierarchical Taxonomy of Psychopathology (HiTOP): A quantitative nosology based on consensus of evidence. *Annual Review of Clinical Psychology*, 17(1), 83–108. <https://doi.org/10.1146/annurev-clinpsy-081219-093304>
- Kruyen, P. M., Emons, W. H. M., & Sijtsma, K. (2013). On the shortcomings of shortened tests: A literature review. *International Journal of Testing*, 13(3), 223–248. <https://doi.org/10.1080/15305058.2012.703734>
- Leite, W. L., Huang, I.-C., & Marcoulides, G. A. (2008). Item selection for the development of short forms of scales using an ant colony optimization algorithm. *Multivariate Behavioral Research*, 43(3), 411–431. <https://doi.org/10.1080/00273170802285743>
- Leite, W. L., Shen, Z., Marcoulides, K., Fisk, C. L., & Harring, J. (2022). Using ant colony optimization for sensitivity analysis in structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 29(1), 47–56. <https://doi.org/10.1080/10705511.2021.1881786>
- Loevinger, J. (1954). The attenuation paradox in test theory. *Psychological Bulletin*, 51(5), 493–504. <https://doi.org/10.1037/h0058543>
- Long, J. S. (1983). *Covariance structure models: An introduction to LISREL*. SAGE.
- MacCallum, R. C., Wegener, D. T., Uchino, B. N., & Fabrigar, L. R. (1993). The problem of equivalent models in applications of covariance structure analysis. *Psychological Bulletin*, 114(1), 185–199. <https://doi.org/10.1037/0033-2909.114.1.185>
- Mansolf, M., & Reise, S. P. (2016). Exploratory bifactor analysis: The Schmid-Leiman orthogonalization and Jennrich-Bentler analytic rotations. *Multivariate Behavioral Research*, 51(5), 698–717. <https://doi.org/10.1080/00273171.2016.1215898>
- Mansolf, M., & Reise, S. P. (2017). When and why the second-order and bifactor models are distinguishable. *Intelligence*, 61, 120–129. <https://doi.org/10.1016/j.intell.2017.01.012>
- Marcoulides, G. A., & Drezner, Z. (1999). Using simulated annealing for model selection in multiple regression analysis. *Multiple Linear Regression Viewpoints*, 25(2), 1–4.
- Marcoulides, G. A., & Drezner, Z. (2001). Specification searches in structural equation modeling with a genetic algorithm. In G. A. Marcoulides & R. E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 247–268). Routledge.
- Marcoulides, G. A., & Drezner, Z. (2003). Model specification searches using ant colony optimization algorithms. *Structural Equation Modeling: A Multidisciplinary Journal*, 10(1), 154–164. [https://doi.org/10.1207/S15328007SEM1001\\_8](https://doi.org/10.1207/S15328007SEM1001_8)

- Marcoulides, G. A., Drezner, Z., & Schumacker, R. E. (1998). Model specification searches in structural equation modeling using tabu search. *Structural Equation Modeling: A Multidisciplinary Journal*, 5(4), 365–376. <https://doi.org/10.1080/10705519809540112>
- Marcoulides, G. A., & Ing, M. (2012). Automated structural equation modeling strategies. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 690–704). The Guilford Press.
- Marcoulides, G. A., & Leite, W. L. (2013). Exploratory data mining algorithms for conducting searches in structural equation modeling. In J. J. McArdle & G. Ritschard (Eds.), *Contemporary issues in exploratory data mining in the behavioral sciences* (pp. 150–171). Routledge.
- Marcoulides, G. A., & Schumacker, R. E. (Eds.). (2001). *New developments and techniques in structural equation modeling*. Routledge.
- Marcoulides, K. M., & Falk, C. F. (2018). Model specification searches in structural equation modeling with R. *Structural Equation Modeling: A Multidisciplinary Journal*, 25(3), 484–491. <https://doi.org/10.1080/10705511.2017.1409074>
- Marsh, H. W., Lüdtke, O., Muthén, B., Asparouhov, T., Morin, A. J. S., Trautwein, U., & Nagengast, B. (2010). A new look at the big five factor structure through exploratory structural equation modeling. *Psychological Assessment*, 22(3), 471–491. <https://doi.org/10.1037/a0019227>
- McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, 37(1), 1–10. <https://doi.org/10.1016/j.intell.2008.08.004>
- Morin, A. J. S., Arens, A. K., & Marsh, H. W. (2016). A bifactor exploratory structural equation modeling framework for the identification of distinct sources of construct-relevant psychometric multidimensionality. *Structural Equation Modeling: A Multidisciplinary Journal*, 23(1), 116–139. <https://doi.org/10.1080/10705511.2014.961800>
- Moshagen, M., Hilbig, B. E., & Zettler, I. (2018). The dark core of personality. *Psychological Review*, 125(5), 656–688. <https://doi.org/10.1037/rev0000111>
- Murohashi, H., & Toyoda, H. (2007). Model specification search using a genetic algorithm with factor reordering for a simple structure factor analysis model. *Japanese Psychological Research*, 49(3), 179–191. <https://doi.org/10.1111/j.1468-5884.2007.00345.x>
- Nelson, L. D., Simmons, J., & Simonsohn, U. (2018). Psychology's renaissance. *Annual Review of Psychology*, 69(1), 511–534. <https://doi.org/10.1146/annurev-psych-122216-011836>
- Olaru, G., Schroeders, U., Hartung, J., & Wilhelm, O. (2019). Ant colony optimization and local weighted structural equation modeling. A tutorial on novel item and person sampling procedures for personality research. *European Journal of Personality*, 33(3), 400–419. <https://doi.org/10.1002/per.2195>
- Olaru, G., Witthöft, M., & Wilhelm, O. (2015). Methods matter: Testing competing models for designing short-scale Big-Five assessments. *Journal of Research in Personality*, 59, 56–68. <https://doi.org/10.1016/j.jrp.2015.09.001>
- Persson, B. N., Kajonius, P. J., & Garcia, D. (2019). Revisiting the structure of the Short Dark Triad. *Assessment*, 26(1), 3–16. <https://doi.org/10.1177/1073191117701192>
- Preacher, K. J., & MacCallum, R. C. (2003). Repairing Tom Swift's electric factor analysis machine. *Understanding Statistics*, 2(1), 13–43. [https://doi.org/10.1207/S15328031US0201\\_02](https://doi.org/10.1207/S15328031US0201_02)
- Preacher, K. J., Zhang, G., Kim, C., & Mels, G. (2013). Choosing the optimal number of factors in exploratory factor analysis: A model selection perspective. *Multivariate Behavioral Research*, 48(1), 28–56. <https://doi.org/10.1080/00273171.2012.710386>

- R Core Team. (2022). *R: A language and environment for statistical computing* [Manual]. <https://www.R-project.org/>
- Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research, 47*(5), 667–696. <https://doi.org/10.1080/00273171.2012.715555>
- Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment, 92*(6), 544–559. <https://doi.org/10.1080/00223891.2010.496477>
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1–36. <https://doi.org/10.18637/jss.v048.i02>
- 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*(1), 95–113. <https://doi.org/10.1177/0013164410387348>
- Schroeders, U., Wilhelm, O., & Olaru, G. (2016a). Meta-heuristics in short scale construction: Ant colony optimization and genetic algorithm. *PLOS ONE, 11*(11), Article e0167110. <https://doi.org/10.1371/journal.pone.0167110>
- Schroeders, U., Wilhelm, O., & Olaru, G. (2016b). The influence of item sampling on sex differences in knowledge tests. *Intelligence, 58*(3), 22–32. <https://doi.org/10.1016/j.intell.2016.06.003>
- Seeley, T. D. (1983). Division of labor between scouts and recruits in honeybee foraging. *Behavioral Ecology and Sociobiology, 12*(3), 253–259. <https://doi.org/10.1007/BF00290778>
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22*(11), 1359–1366. <https://doi.org/10.1177/0956797611417632>
- Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2020). Specification curve analysis. *Nature Human Behaviour, 4*(11), 1208–1214. <https://doi.org/10.1038/s41562-020-0912-z>
- Sørensen, K. (2015). Metaheuristics—the metaphor exposed. *International Transactions in Operational Research, 22*(1), 3–18. <https://doi.org/10.1111/itor.12001>
- Steege, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science, 11*(5), 702–712. <https://doi.org/10.1177/1745691616658637>
- Steger, D., Jankowsky, K., Schroeders, U., & Wilhelm, O. (2022). The road to hell is paved with good intentions: How common practices in scale construction hurt validity. *Assessment*. Advance online publication. <https://doi.org/10.1177/10731911221124846>
- van der Linden, W. J. (2005). *Linear models of optimal test design*. Springer.
- Wagenmakers, E.-J., Wetzels, R., Borsboom, D., van der Maas, H. L. J., & Kievit, R. A. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science, 7*(6), 632–638. <https://doi.org/10.1177/1745691612463078>
- Whittaker, T. A. (2012). Using the modification index and standardized expected parameter change for model modification. *The Journal of Experimental Education, 80*(1), 26–44. <https://doi.org/10.1080/00220973.2010.531299>
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag. <https://ggplot2.tidyverse.org>