A Comparison of Metaheuristic Optimization Algorithms for Scale Short-Form Development

Educational and Psychological
Measurement

2020, Vol. 80(5) 910–931

© The Author(s) 2020

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/0013164420906600

journals.sagepub.com/home/epm

Anthony W. Raborn¹, Walter L. Leite² , and Katerina M. Marcoulides³

Abstract

This study compares automated methods to develop short forms of psychometric scales. Obtaining a short form that has both adequate internal structure and strong validity with respect to relationships with other variables is difficult with traditional methods of short-form development. Metaheuristic algorithms can select items for short forms while optimizing on several validity criteria, such as adequate model fit, composite reliability, and relationship to external variables. Using a Monte Carlo simulation study, this study compared existing implementations of the ant colony optimization, Tabu search, and genetic algorithm to select short forms of scales, as well as a new implementation of the simulated annealing algorithm. Selection of short forms of scales with unidimensional, multidimensional, and bifactor structure were evaluated, with and without model misspecification and/or an external variable. The results showed that when the confirmatory factor analysis model of the full form of the scale was correctly specified or had only minor misspecification, the four algorithms produced short forms with good psychometric qualities that maintained the desired factor structure of the full scale. Major model misspecification resulted in worse performance for all algorithms, but including an external variable only had minor effects on results. The simulated annealing algorithm showed the best overall performance as well as robustness to model misspecification, while the genetic

¹Pasco County Schools, Land O'Lakes, FL, USA

²University of Florida, Gainesville, FL, USA

³University of Minnesota, Minneapolis, MN, USA

Corresponding Author:

Anthony W. Raborn, Pasco County Schools, 7227 Land O' Lakes Blvd. Land O' Lakes, FL 34638, USA.

Email: araborn@pasco.k12.fl.us

algorithm produced short forms with worse fit than the other algorithms under conditions with model misspecification.

Keywords

short-form selection, metaheuristic algorithms, validity, reliability, confirmatory factor analysis, scale development

Applied researchers using psychometric scales often face a dilemma due to limited resources: should they use the full form of a well-established scale with strong validity evidence supporting it, but with a large number of items requiring a substantial amount of time and effort to complete, or should they use a short form of the scale that has not had extensive evidence of validity collected? This issue has generated strong interest in the academic community in scale short forms, and multiple methods have been proposed for scale short-form development (Leite et al., 2008; Olaru et al., 2015). For example, these methods include theoretical or practical justifications for the inclusion or exclusion of items (e.g., Noble et al., 2013), keeping one item from a set of items that are apparently similar or redundant (e.g., Dennis, 2003), obtaining certain criteria for statistical values such as high factor loadings or item correlations (e.g., Byrne & Pachana, 2011; Wester et al., 2012), adding or retaining items that seem to improve measures of reliability and/or dimensionality (e.g., Lim & Chapman, 2013; Veale, 2014). The focus of item selection for short forms therefore tends to be on the internal structure of the newly created form rather than on using external relationships to help build the short form.

Obtaining a short form that has both adequate internal structure and strong validity with respect to relationships with other variables is difficult with traditional methods of short form development. Metaheuristic optimization algorithms (Dréo et al., 2006) have the potential to solve these difficulties because they can simultaneously maximize multiple validity criteria for the developed short forms. This article aims to evaluate multiple automated methods for short-form creation based on metaheuristic optimization algorithms that incorporate criteria based on internal structure and determine which perform best under commonly used scale structures and conditions with and without an external variable.

Metaheuristic Algorithms for Short-Form Selection

There have been some attempts to develop algorithms to derive short forms of scales that (a) maintain the internal structure of the scale in question (e.g., factor structure and/or content balance), (b) have favorable model characteristics such as meeting model fit statistic thresholds, and (c) produce scale scores that have favorable relationships with other variables, including other scales or external variables. For example, Olaru et al. (2015) compared multiple algorithms for the purpose of creating

psychometrically valid short forms of a 99-item Trait Self-Description Inventory scale with various criterion (e.g., maximizing the difference between comparative fit index [CFI] and root mean square error of approximation [RMSEA]) and concluded that, under their study conditions, the ant colony optimization (ACO) algorithm and genetic algorithm (GA) were able to produce statistically appropriate short forms that generalize well to new data. Yarkoni (2010) also examined the use of a GA for the creation of abbreviated versions of a full scale. Marcoulides and Drezner (2004) have shown that a Tabu search (TS) can be used to successfully reduce the number of items loading on factors. Leite et al. (2008) developed and demonstrated an ACO algorithm that selects items for short forms while keeping adequate model fit and maximizing the relationship between the latent variable and external variables. More recently, Browne et al. (2018) produced an automated structural equation modeling (SEM) scale reduction algorithm and purported that it is an effective and efficient method for reducing items during scale development.

While these articles demonstrate the use of some automated scale short-form development strategies and the importance of research in this area, a comparison of automated strategies under different scenarios and measurement settings is not available in the literature. In addition, some commonly used metaheuristic algorithms for combinatorial problems have never been applied to addressing the psychometric short-form development problem. For example, the inaugural example of the simulated annealing (SA) algorithm is with the Traveling Salesman Problem, which has various algorithms attempt to find the shortest path that travels between n cities exactly once (Kirkpatrick et al., 1983), while no psychometric use of SA is apparent in the literature. To address these issues, this article presents a simulation study of scale short-form development utilizing four metaheuristic algorithms: The ACO, GA, TS, and SA. We chose these algorithms because they are the most well-established metaheuristic algorithms in the combinatorial optimization literature (Marti et al., 2018). Flowcharts of these algorithms outlining the process followed to create short forms are shown in Figure 1 and described in detail next.

The ACO algorithm (Colormi et al., 1992) mimics the behavior of ants searching for the shortest path to a food source. As they search, ants leave a pheromone trail and the ants who reach the food source with the shortest path leave the strongest amount of pheromone, which is detected by other ants. After a while, all ants converge to the shortest path. We evaluate an implementation of the ACO algorithm similar to the one presented by Leite et al. (2008) for short-form development that maximizes confirmatory factor analysis (CFA) model fit and relationship between the factors and an external variable. Their implementation of the ACO algorithm attaches sampling weights to items based on pheromone levels, which are used to sample items for a set of candidate short forms of the scales. The pheromone level used by Leite et al. (2008) was a function of model fit and relationships between factors and a single covariate. Researchers using the ACO algorithm have the flexibility of defining the pheromone level according to the characteristics of the short form they want to maximize. The pheromone level is then used to update the sampling

Figure 1. Flowcharts of metaheuristic algorithms for scale short-form selection.

weights for the next round of sampling of candidate sets of short forms. This is repeated until a specified convergence criterion is met, such as the number of iterations without improvement of the solution quality.

The GA (Fraser, 1957; Goldberg, 1989) mimics the biological process of evolution using the model parameters as genes. Similar to Yarkoni (2010), to select scale short forms, the GA generates an initial population of candidate short forms of size 200 and, by evaluating the fitness of each candidate short form through a loss function, selects the best 20% of the short forms and repopulates. Between short-form generations, new short forms are created from mutation (randomly changing the items in each short form) and recombination (two short forms exchanging items retained). After a certain number of iterations (100+), the short form with the best fit according to the loss function is retained as the best solution. The loss function penalizes the fit of the models for every item included; this value needs to be tuned to achieve the correct reduction of items.

The TS algorithm implementation was modified from the presentation given in Marcoulides and Falk (2018) to constrain the solution space to a specified number of items. As applied to short-form development, the TS looks at each set of short forms created by changing one item at a time. The main idea behind the TS procedure is to continually adjust the currently selected best short form by examining other short forms in the neighborhood of the current best short form. If a neighboring short form examined fits better than the current short form, it is selected as the new best fitting short form. If not, the examined neighboring short form is marked “tabu”—placed on a list so that it is not reconsidered until certain criteria are met. For a general description of TS, see Drezner et al. (1999). For this study, the TS started by randomly generating a short form of a predetermined length from a longer form for the first iteration and searched for neighboring local short forms that maintained a predetermined length.

The SA algorithm (Cerny, 1985; Kirkpatrick et al., 1983) is a statistical analog to metallurgic processes of annealing metals (Drezner & Marcoulides, 1999). As applied to scale short-form development, the algorithm begins with a specified short form in which some of the items are randomly changed. The new short form is compared with the starting short form, and the difference in model fit is calculated. At any time, if the new short form has better fit than the current short form, it is selected for use in the next iteration; otherwise, the new short form is selected with probability equal to a function of the difference in model fit and the difference between the current and maximum iterations. The iterations follow a Markov chain because each configuration depends only on the previous configuration. This process continues until the maximum number of iterations is reached and the best short form is returned as the final solution. The SA algorithm is the only metaheuristic algorithm that has been proven to always converge to the global optimum under certain conditions (Dréo et al., 2006).

Research Questions

Investigating the application of the above outlined metaheuristic for the creation of scale short forms, the following research questions were considered:

1. How do the algorithms differ in terms of convergence time, model fit, and reliability of the short form, and ability to exclude problematic items?
2. How do measurement model misspecifications with the full form affect the fit and reliability of the short forms created by the algorithms?
3. Does the inclusion of a covariate affect the model fit and reliability of the short forms and the exclusion of problematic items?
4. Does the performance of the algorithms depend on the factor structure of the scale?

Method

Manipulated Conditions

To examine the above research questions, a Monte Carlo simulation study was conducted using the following population CFA models: (a) the 20-item unidimensional model of the self-deceptive enhancement scale (Leite & Beretvas, 2005), (b) the 24-item three-dimensional model of the teacher efficacy scale (Tschannen-Moran & Hoy, 2001), and (c) the three-factor bifactor model (Splett et al., 2017) of the 27-item Behavior Assessment System for Children, Second Edition Behavioral and Emotional Screening System (Reynolds et al., 2011). These models represent three common CFA models seen in scale development and are good representations of what applied researchers would regularly work with. The covariance structures presented in the cited papers for these scales were used to simulate samples for the target conditions. In each case, the goal was to create a short form that is no more than half the length of the long form; this was chosen since the three-factor teacher efficacy scale already has a short form that is exactly half the length of the full form. Therefore, we sought short forms that were 10 items (one-factor scale) and 12 items (three-factor and bifactor scales) long.

For manipulating the relationship with an external variable, the two levels that were manipulated are (a) no relationship and (b) a moderately large relationship (approximately equal to a path coefficient of 0.6). In conditions with no relationship, the CFA models fit for the short forms only contained the items as endogenous variables. In conditions with a relationship with an external variable, selection of short forms was done with a structural equation model that extended the CFA models by including direct effects from the latent factors to the external variable, which was treated as a distal outcome.

The full-scale model misspecification condition was manipulated in the simulation according to three levels: (a) no misspecification (the baseline models), (b) a minor misspecification in the factor loadings (i.e., population models modified to have six of the items cross-loading on a nuisance factor with a loading of 0.3), and (c) a major misspecification in the factor loadings (i.e., same as [b] but with factor loadings of 0.6 on a nuisance factor).

Software Implementation

The data were simulated in R v3.5.0 (R Core Team, 2018) using the MASS package (Venables & Ripley, 2002). The package uses eigendecomposition to create data from the item means and item covariance matrix; the baseline condition (high reliability, no external variable relationship, no misspecification) used the values provided by the original models as the population values. These values were changed as necessary to create new population models that fit the target simulated conditions, resulting in 15 constructed covariance matrices. The sample size of each condition was set to 500; this is generally considered sufficient for the number of items in each

Table 1. Values of Tuning Parameters of Metaheuristic Algorithms.

Algorithm	Tuning parameter	Form type	Value
ACO	Consecutive steps for convergence	All	20
	Evaporation	All	0.9
	Ants	All	20
	Maximum steps for no improvement	All	1,000
TS	Number of iterations	All	50
	Tabu list size	All	5
SA	Number of iterations	One	500
		Three	1000
		Bifactor	3000
	Maximum item changes per iteration	One	5
		Three	4
		Bifactor	2
	Temperature function	All	Logistic
	Acceptance probability	All	Kirkpatrick
	Restart criteria iterations	All	25
GA	Number of iterations	One	1000
		Three	1000
		Bifactor	1500
	Population size	One	100
		Three	200
		Bifactor	200
	Cost per item	One	0.005
		Three	0.005
		Bifactor	0.00000001

Note. ACO = ant colony optimization; TS = Tabu search; SA = simulated annealing; GA = genetic algorithm.

of the models (Jackson, 2003). For each combination of manipulated conditions, we created 100 data sets.

The scale short-form selection with ACO, TS, and SA was implemented with each imputed data set using the *ShortForm* package (Raborn & Leite, 2018) in *R*, which uses the *lavaan* package (Rosseel, 2012) in the background to fit the one-factor, three-factor, and bifactor CFA models. The scale short-form selection with GA was implemented by modifying the main function of the *GAabbreviate* package (Scrucca & Sahdra, 2016), which runs the *GA* package to implement the GA (Scrucca, 2013). The algorithms were set to maximize the CFI.

The tuning parameters used for each metaheuristic algorithm are shown in Table 1. These tuning parameters were chosen based on previous research (e.g., Leite et al., 2008; Marcoulides & Falk, 2018). The tuning parameters affect the speed at which the algorithms converge onto a solution, the criteria the algorithms use in determining the quality of the short form, and how much exploration of new short

forms is permitted. For each condition, the parameters were chosen to allow for as much direct comparability between each method as possible.

For the current study, we made two modifications to ACO tuning parameters originally used by Leite et al. (2008). We used the CFI as the only fit index within the algorithm, instead of three fit indices, to improve comparability to the TS and SA algorithms, and we required that the algorithm select the same combination of items for 20 consecutive iterations for convergence, instead of converging after no improvement for 1,000 iterations. We then implemented the algorithm with 20 short forms selected at a time and with a pheromone function defined as zero if $CFI > 0.95$ or

$$\phi_j = \frac{\sum_{i=1}^I (\gamma_i)}{I}$$

otherwise, where j indexes the current iteration, i indexes the items included in the best model observed by the algorithm up to the current iteration, and γ is the standardized factor loading of the indicator variables of the scale. The initial pheromone for each item is set to 2 and is updated at the end of each iteration by first multiplying the current pheromone level for each item by the evaporation rate, then adding the value ϕ_j to the items included in the best model observed so far.

The GA from the *GAabbreviate* package was modified to use continuous variables (i.e., factor scores) but otherwise kept the same general implementation. The cost function, $Cost = Ik + \sum_{i=1}^s w_i (1 - R_i^2)$, seeks to minimize the unexplained variance in each scale $i \in 1, \dots, s$ with a penalty I applied to every item k retained across all of the scales in the measure. To ensure that the algorithm reached the target number of items in each scale, we set I to .005 for the one- and three-factor conditions and 0.00000001 for the bifactor conditions. The population size (number of potential candidate models in each iteration) was set to 100 for the one-factor conditions and 200 for the three- and bifactor conditions. Finally, the maximum number of iterations was set to 1,000 for the one- and three-factor conditions and 1,500 for the bifactor conditions. The mutation probability (probability of random changes to candidate models) and crossover probability (probability of paired candidate models to exchange “genes” or item selection indicators) were left at their default values of 0.1 and 0.8, respectively.

The TS algorithm was implemented with a Tabu size of 5 across each condition. This value was used as an intermediate between 2 times the number of latent variables suggested in Marcoulides et al. (1998; 2, 6, and 8 for the one-factor, three-factor, and bifactor conditions) and 10 (Mills et al., 2005). We also specified that the algorithm run for 50 iterations total. Finally, we added an additional step at the beginning of the algorithm that randomly generates a short form that fits the final criteria for that condition.

The SA algorithm was implemented with a logistic temperature (Kirkpatrick et al., 1983). As the success of the SA algorithm can be affected by the selection of the starting temperature, we elected to use this temperature because it allows the

algorithm to broadly search the solution space in the beginning of the search, followed by a period of fairly linear cooling resulting in a steady tightening of the solution space around higher performing solutions, and ending with a more restricted solution space in the coolest final iteration. We allowed the algorithm to run for 500 iterations in the one-factor model conditions, 1,000 iterations in the three-factor model iterations, and 3,000 iterations in the bifactor model iterations. We also restricted the maximum number of changes between iterations to four changes max in the three-factor conditions and two changes in the bifactor model conditions.

Outcomes

A variety of outcome variables can be used to assess the overall performance of the simulations conducted. The outcome variables of the simulations used in this study were the time to converge to a short form, the average level of model fit of the short form, the composite reliability of the short form for each factor, and the number of times the algorithms selected the apparent best items (e.g., those with the highest loadings) in the simulation for the short form.

For this study, the composite reliability (CR) with the one- and three-factor models was calculated as

$$CR_{\omega} = \frac{\left(\sum_{i=1}^I \lambda_i\right)^2}{\left(\sum_{i=1}^I \lambda_i\right)^2 + \sum_{i=1}^I \theta_i} \quad (1)$$

where the items are indexed with i , λ are the standardized factor loadings, and θ are the (standardized) residual variances of the items (McDonald, 1999). For the bifactor model, the composite reliability for the specific factors was calculated as

$$CR_{\omega_s} = \frac{\left(\sum_{i=1}^{I_s} \lambda_i\right)^2}{\left(\sum_{i=1}^{I_s} \lambda_{g_i}\right)^2 + \left(\sum_{i=1}^{I_s} \lambda_i\right)^2 + \sum_{i=1}^{I_s} \theta_i} \quad (2)$$

which uses the factor loadings on both the specific factor λ_i and general factor λ_{g_i} for the items that load on the specific factor I_s . For the general factor, the composite reliability was calculated as

$$CR_{\omega_h} = \frac{\left(\sum_{i=1}^k \lambda_{g_i}\right)^2}{\left(\sum_{i=1}^k \lambda_{g_i}\right)^2 + \sum_{s=1}^J \left(\left(\sum_{i=1}^{I_s} \lambda_{s_i}\right)^2\right) + \sum_{i=1}^k \theta_i} \quad (3)$$

where s indexes the specific factors (Gignac & Watkins, 2013).

The CFI, Tucker-Lewis Index (TLI), and RMSEA were used as the model fit indices, and the cutoff values of CFI > 0.95, TLI > 0.95, and RMSEA < 0.05

were used as indicators of close model fit (Hu & Bentler, 1999). Although these cut-off criteria have been shown to be conservative and sensitive to type of model (Fan & Sivo, 2005, 2007; Leite & Stapleton, 2011; Marsh et al., 2005), they provide an adequate basis to compare across short-form selection methods for the same CFA model.

Results

For each factor model, results for the “Minor Error With External Variable” condition did not have noticeable differences from either of the “Minor Error With No External Variable” conditions, so this condition was dropped from the current study.

One-Factor Model

The time to complete for each algorithm was similar across all conditions, except for GA (see Table 2), which was faster. The time to converge was slightly longer for the ACO, SA, and TS algorithms under the major error with an external relationship condition.

Table 2. Model Fit of Short Forms for the One-Factor Model.

Error/external	Method	Minutes to complete	CFI	TLI	RMSEA
None/no	ACO	2.482	0.976	0.969	0.043
	SA	2.516	0.993	0.992	0.018
	TS	3.933	0.985	0.981	0.028
	GA	0.699	0.975	0.967	0.042
Minor/no	ACO	2.575	0.961	0.950	0.055
	SA	2.572	0.987	0.983	0.027
	TS	3.987	0.977	0.971	0.036
	GA	0.708	0.964	0.953	0.051
Major/no	ACO	2.713	0.943	0.926	0.061
	SA	2.581	0.983	0.978	0.029
	TS	3.956	0.940	0.923	0.061
	GA	0.701	0.850	0.807	0.112
None/yes	ACO	3.059	0.943	0.929	0.058
	SA	2.871	0.981	0.976	0.029
	TS	2.819	0.923	0.903	0.066
	GA	0.702	0.859	0.823	0.106
Major/yes	ACO	3.301	0.942	0.928	0.058
	SA	3.107	0.981	0.976	0.029
	TS	5.162	0.934	0.917	0.060
	GA	0.752	0.855	0.819	0.108

Note. Boldfaced values indicate good model fit. CFI = comparative fit index; TLI = Tucker-Lewis Index; RMSEA = root mean square error of approximation; ACO = ant colony optimization; SA = simulated annealing; TS = Tabu search; GA = genetic algorithm.

The average model fit statistics for both factor models across 100 replications of the analysis for the three conditions is also shown in Table 2, where bolded values indicate close model fit based on the criteria specified before. When there is no error in the original model, each algorithm produced good model fit, but as the error level increased, the model fit decreased. In the major error conditions, only the SA algorithm had model fit greater than the cutoff values used. The inclusion of the external variable reduced the model fit for each of the algorithms such that the conditions with no error and a moderate external relationship had similar fit to the conditions with major error and either no or a moderate external relationship (see Table 2).

Table 3 shows the reliability of the factor scores of the full form of the scale, as well as the reliability of the factor scores of the short forms for each combination of error level and external relationship. As expected, the full form resulted in scores with somewhat higher reliability values than all the short forms. The ACO had the greatest composite reliability for both the no error/no external relationship and minor error/no external relationship conditions, while the GA had the greatest composite reliability for the remaining three conditions but had poor model fit (see Table 2).

The first five rows of Table 4 show the average proportion of iterations the six misspecified items were selected for each condition by each algorithm. As the error condition increased, the ACO and SA algorithms selected the misspecified items less often, and the inclusion of the external variable also resulted in the misspecified items being selected less often. The TS algorithm was unaffected by the error condition but showed a small decrease in the selection of the misspecified items with the inclusion of the external variable. The GA selected the misspecified items more often as error increased and with the inclusion of the external variable.

Three-Factor Model

For all five conditions, the SA and TS algorithm took about twice as long to converge on average as compared with the ACO and GA algorithms (see Table 5). For the three-factor model, the average model fit for the conditions with no external variable can be seen in Table 5. In both the no error and minor error conditions, each of the algorithms selected items with close model fit. In the major error conditions, only the SA algorithm produced short forms with adequate model fit according to all three fit indices. The ACO and TS algorithms had close model fit according to CFI and TLI (ACO) and CFI (TS), while the GA algorithm had poor fit according to all three fit indices.

Including an external variable had little effect on the model fit statistics (see Table 5). With no error, the average model fit was approximately the same between the no external and moderate external conditions, while the average fit somewhat decreased in the major error condition for the moderate external condition as compared with the no external condition. Only the SA had good model fit across all the conditions. The ACO produced short forms with the largest composite reliability for each of the factors in each condition.

Table 3. One-Factor Short-Form Reliability.

Error/external	Method	Reliability
None/no	Full form	0.889
	ACO	0.854
	SA	0.813
	TS	0.806
	GA	0.835
Minor/no	ACO	0.854
	SA	0.812
	TS	0.806
	GA	0.838
	ACO	0.828
Major/no	SA	0.785
	TS	0.799
	GA	0.837
	ACO	0.830
	SA	0.786
None/yes	TS	0.798
	GA	0.848
	ACO	0.828
	SA	0.785
	TS	0.799
Major/yes	GA	0.837
	ACO	0.830
	SA	0.786
	TS	0.798
	GA	0.848

Note. ACO = ant colony optimization; SA = simulated annealing; TS = Tabu search; GA = genetic algorithm.

Table 4. Proportions of Misspecified Items Selected in Each Model.

Model	External	Error	ACO	SA	TS	GA
One Factor	None	None	0.57	0.49	0.45	0.26
		Minor	0.66	0.36	0.40	0.40
		Major	0.30	0.21	0.43	0.46
	Moderate	None	0.20	0.13	0.36	0.39
		Major	0.20	0.14	0.34	0.39
		Major	0.18	0.12	0.36	0.50
Three Factor	None	None	0.46	0.38	0.51	0.48
		Minor	0.42	0.31	0.41	0.54
		Major	0.19	0.12	0.35	0.50
	Moderate	None	0.35	0.39	0.52	0.50
		Major	0.18	0.12	0.36	0.50
		Major	0.18	0.12	0.36	0.50
Bifactor	None	None	0.37	0.49	0.48	0.36
		Minor	0.35	0.47	0.49	0.41
		Major	0.35	0.28	0.42	0.46
	Moderate	None	0.39	0.54	0.50	0.49
		Major	0.28	0.21	0.38	0.50
		Major	0.28	0.21	0.38	0.50

Note: Boldfaced values are the smallest within that condition. ACO = ant colony optimization; SA = simulated annealing; TS = Tabu search; GA = genetic algorithm.

Table 5. Model Fit for Short Forms With Three-Factor Structure.

Error/external	Method	Minutes to complete	CFI	TLI	RMSEA
None/no	ACO	2.967	0.979	0.973	0.043
	SA	6.228	0.991	0.989	0.024
	TS	6.063	0.986	0.982	0.031
	GA	2.705	0.977	0.970	0.044
Minor/no	ACO	3.091	0.978	0.971	0.047
	SA	6.238	0.989	0.985	0.030
	TS	6.264	0.984	0.979	0.037
	GA	2.710	0.974	0.967	0.048
Major/no	ACO	3.285	0.964	0.953	0.054
	SA	6.295	0.990	0.987	0.027
	TS	6.061	0.956	0.943	0.056
	GA	2.695	0.910	0.883	0.087
None/yes	ACO	3.713	0.978	0.971	0.047
	SA	7.040	0.989	0.985	0.030
	TS	7.577	0.984	0.979	0.037
	GA	2.685	0.974	0.967	0.048
Major/yes	ACO	3.387	0.960	0.948	0.060
	SA	6.173	0.984	0.979	0.036
	TS	6.586	0.950	0.935	0.063
	GA	2.282	0.917	0.892	0.087

Note: Boldfaced values indicate good model fit. CFI = comparative fit index; TLI = Tucker-Lewis Index; RMSEA = root mean square error of approximation; ACO = ant colony optimization; SA = simulated annealing; TS = Tabu search; GA = genetic algorithm.

Table 6 shows the composite reliability of the full form of the scale for each factor, as well as the composite reliability of the short forms for each combination of error level and external relationship. As expected, the full form again resulted in scores with higher reliability on each factor than all the short forms. The ACO had the greatest composite reliability in all conditions.

The misspecified item selection proportions are shown in the middle five rows of Table 4. Both the ACO and the SA algorithms selected the misspecified items less often as the error rate increased and with the inclusion of the external relationship. The TS algorithm showed a similar, but smaller, decrease as well, but the GA algorithm had no major change across any of the conditions.

Bifactor Model

With the bifactor model, the GA had the fastest time to converge, and the ACO took about 4 times longer (see Table 7). The TS and SA algorithms had convergence times that were about 10 times that of the GA algorithm. Table 7 also shows the average model fit indices of the bifactor model. The SA and TS algorithms produced short forms with good model fit by each fit index in every condition, while the ACO

Table 6. Three-Factor Short-Form Reliability.

Error/external	Method	Reliability Factor 1	Reliability Factor 2	Reliability Factor 3
None/no	Full form	0.870	0.910	0.900
	ACO	0.788	0.846	0.833
	SA	0.752	0.829	0.813
	TS	0.754	0.828	0.809
	GA	0.763	0.846	0.819
Minor/no	ACO	0.782	0.846	0.830
	SA	0.747	0.829	0.807
	TS	0.744	0.826	0.807
	GA	0.758	0.841	0.820
Major/no	ACO	0.772	0.842	0.820
	SA	0.751	0.829	0.804
	TS	0.733	0.823	0.797
	GA	0.746	0.830	0.804
None/yes	ACO	0.782	0.862	0.845
	SA	0.768	0.834	0.805
	TS	0.766	0.835	0.808
	GA	0.769	0.847	0.821
Major/yes	ACO	0.772	0.842	0.820
	SA	0.751	0.829	0.804
	TS	0.733	0.823	0.797
	GA	0.746	0.830	0.804

Note. ACO = ant colony optimization; SA = simulated annealing; TS = Tabu search; GA = genetic algorithm.

resulted in good model fit by each fit index except for the RMSEA in the major error condition. The GA had good model fit by CFI in the no and minor error conditions only.

Including the external variable tended to reduce model fit. Both the SA and TS algorithms showed slight reductions in model fit across both error conditions but still found short forms with good model fit according to all three statistics. The ACO algorithm maintained approximately the same model fit in both no error conditions but showed an increase in average fit in the major error conditions when comparing the no external to moderate external relationship condition. However, only the CFI and TLI showed good model fit in these conditions (see Table 7).

The reliability of scores of the general factor with the full form and short forms with the bifactor model is shown in Table 8. For the example scale used in this study, the full form produced scores with adequate reliability for the general factor, but for the specific factors, the composite reliability is low. For the short forms, none of the algorithms produced consistently greater reliabilities for every factor in these conditions. The reliability of general factor with the short forms were smaller than the reliability of the full form for all algorithms. The GA performed best for the general factor reliability compared with ACO, SA, and TS. For the specific factors, the TS performed better than the other methods for two out of three

Table 7. Model Fit of Short Forms With Bifactor Model and External Variable.

Error/external	Method	Minutes to complete	CFI	TLI	RMSEA
None/no	ACO	13.054	0.986	0.976	0.041
	SA	37.925	0.995	0.992	0.020
	TS	39.933	0.992	0.986	0.028
	GA	3.577	0.951	0.939	0.074
Minor/no	ACO	12.296	0.982	0.969	0.047
	SA	40.658	0.994	0.990	0.023
	TS	48.200	0.990	0.984	0.031
	GA	3.562	0.954	0.943	0.069
Major/no	ACO	12.875	0.974	0.956	0.055
	SA	39.556	0.993	0.987	0.027
	TS	41.211	0.982	0.969	0.043
	GA	3.606	0.945	0.921	0.079
None/yes	ACO	18.537	0.986	0.976	0.041
	SA	45.815	0.995	0.992	0.020
	TS	49.665	0.992	0.986	0.028
	GA	3.458	0.951	0.939	0.074
Major/yes	ACO	17.989	0.975	0.959	0.051
	SA	42.336	0.991	0.986	0.028
	TS	38.947	0.979	0.965	0.046
	GA	3.546	0.940	0.907	0.080

Note: Boldfaced values indicate good model fit. CFI = comparative fit index; TLI = Tucker-Lewis Index; RMSEA = root mean square error of approximation; ACO = ant colony optimization; SA = simulated annealing; TS = Tabu search; GA = genetic algorithm.

factors. Surprisingly, the TS produced scores with higher reliability than the full form for Factor 3 and for Factor 1 in the no error and major error with an external variable conditions, and all algorithms produced higher reliability than the full form for Factor 2. These results are limited in that the scale produced scores with low reliability. Using a different scale that results in higher reliability of scores of the full form for all factors might have produced different results with respect to the comparison of algorithms.

The last five rows of Table 4 show the misspecified item selection proportions for the bifactor conditions. In the no external variable conditions, there were very little differences between the item selection proportions between the no error and the minor error conditions for each algorithm, and only the SA and TS algorithms showed a decrease in the proportion of misspecified item selection when comparing the minor error to the major error condition. In the external variable conditions, the ACO, SA, and TS algorithms all showed a decrease in misspecified item selection, while the GA stayed relatively constant.

Table 8. Bifactor Model Short-Form Reliability.

Error/external	Method	Reliability bifactor	Reliability Factor 1	Reliability Factor 2	Reliability Factor 3
None/no	Full form	0.715	0.304	0.114	0.319
	ACO	0.661	0.067	0.115	0.062
	SA	0.641	0.061	0.133	0.055
	TS	0.659	0.149	0.124	0.433
	GA	0.669	0.055	0.118	0.049
Minor/no	ACO	0.659	0.067	0.115	0.062
	SA	0.624	0.063	0.141	0.062
	TS	0.657	0.186	0.125	0.384
	GA	0.641	0.057	0.129	0.060
Major/no	ACO	0.643	0.066	0.124	0.067
	SA	0.572	0.066	0.158	0.087
	TS	0.540	0.147	0.162	0.538
	GA	0.582	0.067	0.146	0.081
None/yes	ACO	0.641	0.073	0.115	0.066
	SA	0.657	0.052	0.124	0.055
	TS	0.664	0.437	0.122	0.308
	GA	0.678	0.044	0.121	0.047
Major/yes	ACO	0.615	0.069	0.131	0.070
	SA	0.586	0.064	0.147	0.078
	TS	0.498	0.371	0.172	0.365
	GA	0.374	0.147	0.199	0.116

Note: Boldfaced values also had acceptable model fit on at least two fit statistics. ACO = ant colony optimization; SA = simulated annealing; TS = Tabu search; GA = genetic algorithm.

Discussion

In general, the examined algorithms produced short forms with close model fit in all cases with no error, with two exceptions: the GA under the bifactor and each algorithm except SA under the one-factor model with an external variable. Therefore, when the original scale is correctly specified, the results showed that the algorithms are likely to produce short forms with good psychometric qualities that maintain the desired factor structure of the full scale.

The ACO, TS, and GA each had problems maintaining close model fit for the short forms with increasing error, though this was alleviated by increasing the factor structure's complexity. Including an external variable into the process generally had a small negative effect on average model fit, but the effect was never enough to cross the model fit thresholds. Overall, the SA provided short forms with the best average model fit in every single condition, while the ACO seemed to result in higher composite reliability on average for each of the factors with one- and three-factor models. Given that the difference in composite reliability between the ACO and SA algorithms was about .05 or less on average, the practical difference in reliability between these methods may be outweighed by the difference model fit of the resulting short

forms. Therefore, the current results lead to recommending the SA as the preferable metaheuristic algorithm for automated short-form selection.

The reasons for the success of SA in the studied circumstances are not entirely clear. However, because SA simulates a Metropolis-Hastings Monte Carlo search procedure of the constrained solution space of the short-form creation process, it may be particularly suited to finding a nearly optimal solution that approximates the global solution of the sample (i.e., a short form with the best possible model fit with the current sample). Changes to the optimization function (e.g., using a function of model fit and scale reliability) could change the efficacy of this algorithm for the better, but further research into this possibility is clearly needed.

The ACO, on the other hand, used the factor loadings in the pheromone function, which determines item inclusion probabilities. This explains its better performance regarding overall reliability because reliability calculations in factor analytic models are based in part on factor loadings (see Equations 1-3). However, the relatively less important role of overall model fit with ACO (as compared with SA) results in lower average model fit. This algorithm may achieve better results overall if the minimum model fit of the intermediate short forms was increased, if additional fit statistics were included in the model fit requirements, or if the pheromone function utilized fit index values differently than the current implementation.

Under many of the conditions, the TS algorithm took substantially longer than the other three algorithms to select the short form. This may be due to the definition of the neighborhood in this study. Reducing the size of the neighborhood will inevitably reduce the convergence time. However, if the defined neighborhood is too small, the optimal short form is unlikely to be selected. As a solution, the TS algorithm may benefit from adopting a random neighborhood selection approach as in the hybrid TS algorithm proposed by Marcoulides (2018). This approach may reduce convergence time while increasing the “dispersion of the search over the entire solution set” (Marcoulides, 2018).

In contrast to past results that have advocated for the use of GA in the construction of short scales (e.g., Eisenbarth et al., 2015; Sahdra et al., 2016; Yarkoni, 2010), in this study, the GA performed the worst overall, particularly when the original form had misspecification. This is somewhat expected given that the particular implementation of the algorithm has no way to take these specification errors into account directly. Modifying the form of the algorithm to utilize factor analytic methods, instead of linear regression methods, should lead to better performance as the short-form creation process would then be better aligned to the formulation of the full form.

Conclusion

This study provides useful information to applied researchers about the benefits and drawbacks of utilizing these four metaheuristic algorithms for scale short-form development in some common scenarios in educational and psychological research. This

will allow for easier creation of psychometrically sound short forms with stronger evidence of validity, especially as compared with creating short forms manually. Furthermore, if the short form is developed from a full form of a scale with desirable psychometric properties, such as high factor loadings and high composite reliability, the algorithms should be able to identify a short form with psychometric properties close to the full form. However, if the full form has some undesirable properties, such as some items with low factor loadings and high measurement error, the algorithms will be able to select only the best items from the full form, providing a short form with better psychometric properties than the full form.

Among the metaheuristic algorithms investigated, our research shows that the SA algorithm is most robust to misspecification of the full form, and therefore, we recommend it for applied researchers interested in the development of short forms. However, guidelines for selecting tuning parameters of metaheuristic algorithms for short-form development have not been established, and applied researchers have to rely mostly on trial and error. Although the tuning parameters we provide in Table 1 can be used as an initial guide for trying these algorithms with real data, it is evident that additional research is needed on improving ways to best select these parameters.

A limitation of the current study is that the metaheuristic algorithms were implemented based on the CFI as the only model fit criterion to simplify the comparison of algorithms, but they could be used to maximize a function of multiple fit indices, composite reliability, and relationship with multiple external variables. These external variables can be observed or latent and could be predictors or outcomes of the factors measured by the scale of interest. Additionally, the model evaluated during the selection of short form is not restrained to a CFA, it can be any arbitrarily complex SEM. Our study did not explore the performance of metaheuristic algorithms with SEM but, instead, showed that the model fit results for multiple fit indices and composite reliability were robust to misspecification, presence of an external variable, and model types despite the fact that only one fit index was directly used in the algorithms.

In the current study, we evaluated short forms based on a single validation sample. In the literature, the most common suggestion for validating new short forms is to have two samples, where the first sample is used as a “training sample” to select items for the short form, and the second sample is used as a “testing sample” to validate the short form. This allows for the resulting short forms to be tested on a new sample that was not used to create them, potentially validating the results across new samples. When obtaining two samples is unfeasible, single-sample cross-validation methods, such as k -fold cross validation, may be used to emulate a second validation sample and provide an estimate of the generalizability of the competing short forms to a new sample. To the best of our knowledge, these cross-validation methods are not often used in the area of psychometric short-form creation and serve as a potential area of future research (see Arlot & Celisse, 2010, for a survey of various cross-validation methods and some theoretical evidence supporting their use). While this article explored four commonly used metaheuristic algorithms, there are of course

many other algorithms that could also be examined in future studies (e.g., particle swarm optimization, memetic algorithm, stochastic hill climbing, and many others; Brownlee, 2011; Marti et al., 2018). Additionally, this study only examined sample sizes with 500 observations; therefore, it is unclear whether sample size may affect the algorithms' capabilities of finding a solution. Finally, we only used one empirical example for each full-form model type. Although unlikely, it is possible that other more complex empirical examples may influence the efficacy of the tested algorithms, particularly for the bifactor model.

Future research could compare these metaheuristic algorithms for the short-form creation problem alongside other methods on real data paired with both CFA and SEM. Additionally, these metaheuristic algorithms could potentially be used in virtual learning environments (Aleven et al., 2011; Huggins-Manley et al., 2019) to select a set of items from an item pool for formative assessment. Although the implementation used in this study with the *ShortForm* package of *R* is not fast enough for real-time formative assessment, implementation in other programming languages (e.g., C++) could provide fast convergence times.

Declaration of Conflicting Interests

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

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Walter L. Leite <https://orcid.org/0000-0001-7655-5668>

References

- Aleven, V., Beal, C. R., & Graesser, A. C. (2013). Introduction to the special issue on advanced learning technologies. *Journal of Educational Psychology, 105*(4), 929-931. <https://doi.org/10.1037/a0034155>
- Arlot, S., & Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys, 4*(2010), 40-79. <https://doi.org/10.1214/09-SS054>
- Browne, M., Rockloff, M., & Rawat, V. (2018). An SEM algorithm for scale reduction incorporating evaluation of multiple psychometric criteria. *Sociological Methods & Research, 47*(4), 812-836. <https://doi.org/10.1177/0049124116661580>
- Brownlee, J. (2011). *Clever algorithms: nature-inspired programming recipes*. <http://www.cleveralgorithms.com/nature-inspired/index.html>
- Byrne, G. J., & Pachana, N. A. (2011). Development and validation of a short form of the Geriatric Anxiety Inventory—the GAI-SF. *International Psychogeriatrics, 23*(1), 125-131. <https://doi.org/10.1017/S1041610210001237>

- Cerny, 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>
- Colormi, A., Dorigo, M., & Maniezzo, V. (1992). Distributed optimization by ant colonies. In F. Varela & P. Bourguine (Eds.), *Proceedings of ECAL 91: First European Conference on Artificial Life*. Elsevier.
- Dennis, C. L. (2003). The breastfeeding self-efficacy scale: Psychometric assessment of the short form. *Journal of Obstetric, Gynecologic & Neonatal Nursing*, 32(6), 734-744. <https://doi.org/10.1177/0884217503258459>
- Dréo, J., Pétrowski, A., Siarry, P., & Taillard, E. (2006). *Metaheuristics for hard optimization*. Springer.
- Drezner, Z., & Marcoulides, G. A. (1999). Using simulated annealing for selection in multiple regression analysis. *Multiple Linear Regression Viewpoints*, 25, 1-4.
- Drezner, Z., Marcoulides, G. A., & Salhi, S. (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>
- Eisenbarth, H., Lilienfeld, S. O., & Yarkoni, T. (2015). Using a genetic algorithm to abbreviate the Psychopathic Personality Inventory-Revised (PPI-R). *Psychological Assessment*, 27(1), 194-202. <https://doi.org/10.1037/pas0000032>
- Fan, X., & Sivo, S. A. (2005). Sensitivity of fit indexes to misspecified structural or measurement model components: Rationale of two-index strategy revisited. *Structural Equation Modeling*, 12(3), 343-367. https://doi.org/10.1207/s15328007sem1203_1
- Fan, X., & Sivo, S. A. (2007). Sensitivity of fit indices to model misspecification and model types. *Multivariate Behavioral Research*, 42(3), 509-529. <https://doi.org/10.1080/00273170701382864>
- Fraser, A. S. (1957). Simulation of genetic systems by automatic digital computers. *Australian Journal of Biological Sciences*, 10(4), 484-491. <https://doi.org/10.1071/B19570484>
- Gignac, G. E., & Watkins, M. W. (2013). Bifactor modeling and the estimation of model-based reliability in the WAIS-IV. *Multivariate Behavioral Research*, 48(5), 639-662. <https://doi.org/10.1080/00273171.2013.804398>
- Goldberg, D. E. (1989). *Genetic algorithms in search, optimization and machine learning*. Addison-Wesley.
- Hu, L. T., & 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>
- Huggins-Manley, A. C., Beal, C. R., D'Mello, S. K., Leite, W. L., Cetin-Berber, D. D., Kim, D., & McNamara, D. S. (2019). A commentary on construct validity when using operational virtual learning environment data in effectiveness studies. *Journal of Research on Educational Effectiveness*. <https://doi.org/10.1080/19345747.2019.1639869>
- Jackson, D. L. (2003). Revisiting sample size and number of parameter estimates: Some support for the N:q Hypothesis. *Structural Equation Modeling: A Multidisciplinary Journal*, 10(1), 128-141. https://doi.org/10.1207/S15328007SEM1001_6
- Kirkpatrick, S., Gelatt, C. D., & Vecchi, M. P. (1983). Optimization by simulated annealing. *Science*, 220(4598), 671-680. <https://doi.org/10.1126/science.220.4598.671>
- Leite, W. L., & Beretvas, S. N. (2005). Validation of scores on the Marlowe-Crowne social desirability scale and the balanced inventory of desirable responding. *Educational and Psychological Measurement*, 65(1), 140-154. <https://doi.org/10.1177/0013164404267285>

- 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., & Stapleton, L. (2011). Detecting growth shape misspecifications in latent growth models: An evaluation of fit indices. *Journal of Experimental Education*, 79(4), 361-381. <https://doi.org/10.1080/00220973.2010.509369>
- Lim, S. Y., & Chapman, E. (2013). Development of a short form of the attitudes toward mathematics inventory. *Educational Studies in Mathematics*, 82(1), 145-164. <https://doi.org/10.1007/s10649-012-9414-x>
- Marcoulides, K. M. (2018). Automated latent growth curve model fitting: A segmentation and knot selection approach. *Structural Equation Modeling*, 25(5), 687-699. <https://doi.org/10.1080/10705511.2018.1424548>
- Marcoulides, G. A., & Drezner, Z. (2004). Tabu search variable selection with resource constraints. *Communications in Statistics: Simulation & Computation*, 33(2), 355-362. <https://doi.org/10.1081/SAC-120037240>
- Marcoulides, G. A., Drezner, Z., & Schumacker, R. E. (1998). Model specification searches in structural equation modeling using tabu search. *Structural Equation Modeling*, 5(4), 365-376. <https://doi.org/10.1080/10705519809540112>
- Marcoulides, K. M., & Falk, C. (2018). Model specification searches in structural equation modeling with R. *Structural Equation Modeling*, 25(3), 484-491. <https://doi.org/10.1080/10705511.2017.1409074>
- Marsh, H. W., Hau, K., & Grayson, D. (2005). Goodness of fit in structural equation models. In A. Maydeu-Olivares & J. J. McArdle (Eds.), *Contemporary psychometrics: A festschrift for Roderick P. McDonald* (pp. 275-340). Lawrence Erlbaum.
- Marti, R., Pardalos, P. M., & Resende, M. G. C. (2018). *Handbook of heuristics*. Springer.
- Mills, J. D., Olejnik, S. F., & Marcoulides, G. A. (2005). The tabu search procedure: An alternative to the variable selection methods. *Multivariate Behavioral Research*, 40(3), 351-371. https://doi.org/10.1207/s15327906mbr4003_4
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Lawrence Erlbaum.
- Noble, W., Jensen, N. S., Naylor, G., Bhullar, N., & Akeroyd, M. A. (2013). A short form of the Speech, Spatial and Qualities of Hearing scale suitable for clinical use: The SSQ12. *International Journal of Audiology*, 52(6), 409-412. <https://doi.org/10.3109/14992027.2013.781278>
- 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>
- R Core Team. (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org/>
- Raborn, A. W., & Leite, W. L. (2018). ShortForm: An R package to select scale short forms with the ant colony optimization algorithm. *Applied Psychological Measurement*, 42(6), 516-517. <https://doi.org/10.1177/0146621617752993>
- Reynolds, C. R., Kamphaus, R. W., & Vannest, K. J. (2011). Behavior assessment system for children (BASC). In *Encyclopedia of clinical neuropsychology* (pp. 366-371). Springer. https://doi.org/10.1007/978-0-387-79948-3_1524
- 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>

- Sahdra, B. K., Ciarroch, J., Parker, P., & Scrunna, L. (2016). Using genetic algorithms in a large nationally representative American sample to abbreviate the Multidimensional Experiential Avoidance Questionnaire. *Frontiers in Psychology*, 7, 189. <https://doi.org/10.3389/fpsyg.2016.00189>
- Scrucca, L. (2013). GA: A package for genetic algorithms in R. *Journal of Statistical Software*, 53(4), 1-37. <https://doi.org/10.18637/jss.v053.i04>
- Scrucca, L., & Sahdra, B. K. (2016). *Package 'GAabbreviate'*. <https://cran.r-project.org/web/packages/GAabbreviate/GAabbreviate.pdf>
- Splett, J. W., Raborn, A., Lane, K. L., Binney, A. J., & Chafouleas, S. M. (2017). Factor analytic replication and model comparison of the BASC-2 Behavioral and Emotional Screening System. *Psychological Assessment*, 29(12), 1543-1549. <https://doi.org/10.1037/pas0000458>
- Tschannen-Moran, M., & Hoy, A. W. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education*, 17(7), 783-805. [https://doi.org/10.1016/S0742-051X\(01\)00036-1](https://doi.org/10.1016/S0742-051X(01)00036-1)
- Veale, J. F. (2014). Edinburgh Handedness Inventory–Short Form: A revised version based on confirmatory factor analysis. *Laterality: Asymmetries of Body, Brain and Cognition*, 19(2), 164-177. <https://doi.org/10.1080/1357650X.2013.783045>
- Venables, W. N., & Ripley, B. D. (2002). Generalized linear models. In *Modern applied statistics with S* (pp. 183-206). Springer Science & Business Media. <https://doi.org/10.1007/978-0-387-21706-2>
- Wester, S. R., Vogel, D. L., O'Neil, J. M., & Danforth, L. (2012). Development and evaluation of the Gender Role Conflict Scale Short Form (GRCS-SF). *Psychology of Men & Masculinity*, 13(2), 199. <https://doi.org/10.1037/a0025550>
- Yarkoni, T. (2010). The abbreviation of personality, or how to measure 200 personality scales with 200 items. *Journal of Research in Personality*, 44(2), 180-198. <https://doi.org/10.1016/j.jrp.2010.01.002>