Poster

TITLE

Genes and Ants: Meta-Heuristic Algorithms for Scale Length Optimization

ABSTRACT

Conventional scale reduction strategies often fail to consider multiple psychometric criteria. Computer science offers algorithms for exploring item combinations to identify optimal solutions. The current study compares the psychometrics of multiple shortened scales across three algorithms: removing low factor loadings (i.e., conventional), genetic algorithms, and ant colony optimization.

PRESS PARAGRAPH  
Many organizations use assessments to evaluate, track, and improve the quality of their human capital. However, in many cases, accurate measures of complex attitudes and competencies are prohibitively long leading to a dilemma: do we shorten the measure and sacrifice instrument quality or preserve longer measures to quantify only a few select areas? Advances in Big Data promise efficient solutions for optimizing the tradeoff between length and quality. The current study evaluates the results of using multiple algorithms to shorten multiple scales.

WORD COUNT

2,999

Since the early days of test development, controversy has surrounded the use of shortforms as veridical reflections of their longer parents (Cronbach, 1954), which can be traced to opposing philosophical views on the conduct of quantitative inquiry (Haig, 2013). Purists argue full forms have been constructed over multiple iterations such that every item contributes to test reliability, while providing sufficient construct coverage. Therefore, shortening a measure erodes the cumulative nomological network upon which a test has been built and leads to inflated Type I and Type II error rates (Credé, Harms, Niehorster, & Gaye-Valentine, 2012). From the pragmatic view, the length and redundancy of many psychological instruments places undue burden on participants and should shave away all but what is necessary to capture a construct’s “core” to improve efficiency and latitude in testing multivariate hypotheses.

The current paper treats the dual tension between quality and efficiency as a problem of optimization. Rather than competing pros and cons, the focus of scale abbreviation should be identifying an ideal sub-set given trade-offs across competing criteria. In this sense, a variety of optimization algorithms can be used to identify item sub-sets from longer scales that maximizes the joint distribution of bandwidth *and* fidelity. The current study compares two such algorithms – one based on genetic evolution and the other on ant colony foraging – and their performance in relation to a conventional strategy for scale reduction.

**Optimization Problems in Scale Development**

An optimization problem asks to find either a minimum or maximum value of a given function. Through the *optimization method,* the required solution is found via a search process performing in a combinatorial space of all candidate solutions. Humans also use heuristics to solve problems, such as item-total correlations, maximum factor loadings, or validity coefficients to identify the best performing items from a larger pool. For numerous reasons, however, such strategies are often applied in an unsystematic and sub-optimal manner (Yarkoni, 2010).

For one, scale reduction is often framed as a series of successive stages (Clark & Watson, 1995). One might first factor analyze the full form and retain a sub-set of high loading items, next maximize internal consistency, and then proceed through validation. Prematurely eliminating an item due to a low factor loading in a *developed test,* however, may overlook reasons for its initial retention (e.g., criterion or content validity). Because a small fraction of possible item sub-sets is ultimately considered, sequential approaches are not guaranteed to achieve optimality (Yarkoni, 2010). Additionally, due to limited cognitive resources, researchers often use ad hoc strategies or naïve rules in a heuristic fashion (Schroeders, Wilhelm, & Olaru, 2016), which can often emphasize a certain psychometric criterion while forgoing anothe. Finally, the labor required to abbreviate extensive instruments may deter such endeavors (Yarkoni, 2010). Enumerating and testing all possible item combinations in multidimensional inventories can become prohibitively large. For instance, reducing a 50-item measure to a 15-item short form would require comparing 2,250,829,575,120 separate possible models (= ).

**Meta-Heuristics**

To address the limitations above, *meta-heuristics* have been advanced to make large searches manageable by exploring only sensible parts of a solution space within a reasonable time frame. Rather than search *every* possible item combination, meta-heuristics draw from natural processes (e.g., genes, ants), which use flexible but thorough strategies to guide discovery of *approximate* solutions. Marcoulides and Ing (2012) describe these as automated, yet “intelligent,” specification search methods that try to optimize a variety of psychometric criteria compared with “dumb” specification searches (e.g., factor loadings > .40).

Two popular meta-heuristics are Genetic Algorithms (GA) and Ant Colony Optimization (ACO). GAs evaluate models through successive improvements, similar to how genetic material evolves over generations. GAs rely on Darwinian principles of selection, crossover, and mutation to evolve a solution to the problem of short-form development. If new short-forms are better than the predecessor as the process evolves, the latter is replaced (see Yarkoni, 2010 for details). ACOaims to maximize fit by converging on the correct model in ways that mimic how ants forage for food by accumulating pheromones on the shortest routes. In other words, a set of randomly formed short-scales are initially chosen, a structural equation model is estimated and evaluated with respect to an optimization function for each short-form, more attractive solutions direct the weighting of future item searches, and penalties are imposed to prevent premature convergence at early stages of exploration (see Leite, Huang, & Marcoulides, 2008 for details).

Both heuristics cycle through item combinations, update given an optimization function, allow stochastic elements for exploration, and converge on a sub-set that maximizes user-specified criteria. They diverge, however, in the number of criteria considered, strategy for composing item sub-sets, and duration to completion.

**Current Study**

Using meta-heuristics in scale shortening (Schroeders et al., 2016), we compare the ability of three approaches to produce 5-, 10-, and 15-item versions of each of the seven Hogan Personal Inventory scales (HPI; Hogan & Hogan, 1995) that parallels the dimensionality, reliability, discriminability, and validity of their full counterparts. The HPI is a personality assessment that is frequently used in selection and development. However, it is relatively long and there are no published short-forms. Furthermore, the HPI derivate in the International Item Pool has two primary limitations. The equivalence between forms is questionable as content is not directly derived from the HPI item pool and the IPIP does not present psychometric properties of short-forms abbreviated to varying degrees.

**Methods**

**Sample**

Data used for this study were part of the Hogan Global normative dataset including a total of 145,792 employees who completed the HPI for developmental or selection purposes. To maximize data integrity, we included only cases with full responses and demographics (*n* = 70,967). Further, we removed inattentive cases based on the following criteria: Long-string index of 30 or more consecutive responses (*n* = 1,437) and a within-person correlation coefficients less than .15 (*n* = 5,784) using a psychometric synonym index of over 50 item pairs (*r* > .40). Accounting for partial overlap in carelessness indices, the final sample was 63,805 participants (59% male) with a mean age of 34.62 years (*SD* = 9.98).

**Training and Validation Samples.** To avoid overfitting, we applied an 80-20 random split of the sample into a training sample used to build the short-forms (*n* = 51,044) and a validation sample (*n* = 12,761) used to evaluate the newly derived forms’ performance. In creating these samples, we balance two goals: On one hand, we wanted to maximize the sample size for deriving the short-forms because machine learning algorithms generally improve as more training observations are included. On the other hand, we wanted to retain a sufficiently large validation sample to provide unbiased estimates of how well the optimization worked.

**Measures**

The HPI is an occupational measure of the Five-Factor Model that divides Extraversion and Openness into two sub-facets (Hogan & Hogan, 1995). The latest version of the HPI has 239 items, 41 homogeneous item composites (i.e., subscales), and seven primary scales. In order to assess validity, we analyzed correlations of short forms with the Hogan reliability, sales, and managerial occupational scales. The three scales are empirically derived compounds of other HPI items (hen are included in all samples), designed to maximize prediction of diverging criteria (e.g., counterproductive work behaviors, leadership, objective revenue), and validated for use in personnel selection. The reliability scale is a broad measure of organizational effectiveness focusing on the identification of individuals who are “honest, dependable, and responsive to supervision,” (Hogan & Hogan, 1995); whereas, the sales and managerial occupational scales were designed to be valid predictors of success in their respective focuses.

**Analyses**

All data preparation and analyses were conducted with *R 3.3.1*; CFA models were estimated with the R package *lavaan* (Rosseel, 2012). The ACO script was adopted from the script provided by Schroeders et al. (2016). The GA was implemented by the R package *GAabbreviate* (Scrucca & Sahdra, 2015) which applies the procedure proposed by Yarkoni (2010). All item-selection procedures were separately applied to each HPI scale. The script for implementing a majority of analyses will be made available in a Github repository. Following Schroeders et al. (2016), all solutions were compared along four criteria: (1) Unidimensional model fit (as indicated by CFI and RMSEA), (2) reliability as indicated by William’s *ω*, (3) item endorsement (i.e., difficulty), and (4) external validation.

**Stepwise confirmatory factor analysis (SCFA).** For comparison, we used a conventional approach to item selection based on stepwise removal of items with small factor loadings. This was performed by fitting a unidimensional model for each HPI scale in *lavaan*, removing the lowest item loading, re-estimating the measurement model on the reduced item set, and repeating the process until the predefined number of items was reached. The strategy retains items which seem most central to the factor and homogenous with respect to average inter-item correlations. All CFAs were estimates using a weighted least squares mean-adjusted (WLSM) estimator with all binary item responses classified as ordered responses.

**Genetic Algorithms.** The GA aims to reduce item redundancy by maximizing the square correlation between the total scores of the abbreviated and full scales. The item loss function is a sum of two quantities: (a) An item cost, which increases in direct proportion to the number of items retained by the abbreviated measure, and (b) a variance cost, which increases in proportion to the amount of unexplained variance in the original measure. Formally, the loss function for a single scale can be expressed as:

Where *I* represents a user-specified fixed item cost, *k* represents the number of items retained by the GA (in any given iteration), and *R2* is the amount of variance in in the full scale that can be explained by a linear combination of individual items.

**Ant Colony Optimization***.*ACO incorporates *multiple* distinct criteria in defining the optimization function. Following Schroeder et al. (2016), we address four psychometric criteria: (a) Overall model fit to test for unidimensionality, (b) reliability, (c) item endorsement, and (d) prediction of three Hogan occupational scales. In all cases, model parameters are logit transformed in a way that psychometrically superior item are more heavily preferred (Schroder et al., 2016). The logistic functions place all parameters on a standard scale, accentuates small differences at key cut points (e.g., RMSEA = .05), and maps values to range from 0 to 1.

First, goodness-of-fit was determined using a combination of the incremental fit index of CFA and absolute fit index of RMSEA. We modified the parameters of Schroeders et al. (2016) by placing the CFI inflection point at .90 with a wider logistic curve given the heterogeneous HPI content and often sub-optimal fit indices of most complex personality inventories (Marsh, Morin, Parker, & Kaur, 2014).

The RMSEA cutoff value was retained as indicating good model fit at a value of .05.

Both model fit indicators were averaged into a single model fit index:

Second, reliability was assessed using William’s ω. Despite its ubiquity, Cronbach’s α is based on the assumption that a single factor underlies the scale *and* that all items are equally good indicators of this latent variable (Schmitt, 1996). If the factor loadings are not equal, coefficient omega is a more appropriate estimate of reliability and will always equal or exceed α for a unidimensional scale. Following conventional guidelines, we set a coefficient omega of .70 or higher, which is ideal for item scale creation.

The third criterion assessed refers to the sensitivity of the measure. Selection specialists prefer measures which maximally separate applicants across all levels of a continuum in order to make decisions. In terms of personality, this represents *endorsement* or difficulty rates that match a normal distribution. We specified a quadratic term to cover a broad range of items with mean endorsement rates close to .50. This helps build a scale that captures a normalized trait distribution and prioritizes items which yield maximal information about respondents.

Finally, we assessed the pattern of correlations between the short and long versions with respect to targeted covariates in terms of similarity (e.g., reliability, sales, and managerial occupational scales). The correlation matrices of the short- and long-form would be identical. Accounting for standard errors, a maximum of differences in the correlations | .05 | or below was considered a good representation of the long-form:

The overall optimization function was treated as an equally weighted, additive function of the four criteria:

**Results**

Figure 1 presents a correlogram depicting the correlations between full and short-forms using different methods. There is generally large convergence within abridged versions of the same scale as well as between all abridged versions and their full counterparts. Likewise, the magnitude of the correlations between dissimilar HPI constructs using similar and different methods were generally small supporting discriminant validity. As expected, varying the length of the instrument exerts predictable effects on the brevity-fidelity tradeoff with longer scales retaining more of the variance in the original HPI scales.

Table 1 presents findings on model fit. ACO invariably produced scales with stronger structural validity. More impressive, ACO identified extremely short scales (5-item) with exceptionally good fit across all HPI scales (CFI range = .987 to .999; RMSEA range = .016 to .025). The GA produced slightly weaker, but still acceptable fit indices across the shorter scales (5- and 10-item). It worsened with increased length, which is to be expected as the HPI is a fairly heterogeneous measure and the GA attempts to mirror the full scale as closely as possible. With few exceptions, the SCFA performed worse relative to both meta-heuristics.

Figure 2 provides the loadings distributions. On average, SFCA had the highest loadings (*M* = .64, *SD* = .08), followed by ACO (*M* = .54, *SD* = .07), with the GA (*M* = .49, *SD* = .07) mirroring the same distribution of the original scale (*M* = .49, *SD* = .07). This is expected as SFCA relies solely on discarding items with weak loadings, whereas the GA attempts to reproduce the original scale. These results translate into higher omegas for SFCA (*M*ω = .87) relative to both ACO (*M*ω = .79) and the GA (*M*ω = .73), all of which are lower than the average reliability for the full forms (*M*ω = .88). Nine GA short-forms had omega coefficients falling below .70, whereas ACO short-forms only had two. These results suggest SFCA matches the internal consistency of the long-forms, the GA diminishes it, and ACO finds a compromise.

To determine how shortening affects the discriminating power of the measure, the distribution of item difficulties is plotted in Figure 3. All three strategies do a comparably good job of covering the same range of item endorsements in the full scales. Average endorsements rates varied between .68 and .73. The effect of algorithm on item sensitivity was miniscule (GA: *M* = .69;ACO: *M* = .71) showing only slightly more moderate endorsement rates compared to the full scales (*M* = .73). The “yes” endorsement for most HPI items may restrict the meta-heuristics’ ability to exploit gains in this particular criterion.

Finally, Table 2 presents findings on the capacity of all algorithms to recapture HPI’s associations with external variables. The first three columns are convergent correlations between abbreviated and original versions (*min r* = .56, *max r =* .97), with effects being largest for the GA (*M* = .89) and smallest for SCFA (*M* =.81). In all cases, the GA had the largest convergent correlations and, for a few instances, the discrepancy was stark (e.g., 5-item Prudence). This is sensible given the GA retains items which share maximal overlap with the full scale. The remaining columns present the differences between the abbreviated and full HPI scales correlations with the three occupational scales, where negative numbers represent a loss in validity of the shortened version. The largest deviation occurred for ACO and SCFA algorithms, especially for shorter scales [5 items: *max*(Δ*r*) = -.36; 10 items: *max*(Δ*r*) = -.21; 15 items: *max*(Δ*r*) = -.23]. Across meta-heuristics, the average validity degradation was small (*M*Δ*r*.SCFA = -.05; *M*Δ*r*.GA = -.02; *M*Δ*r*.ACO = -.04) with the greatest variability in deviation occurring for SCFA (*SD*Δ*r*.SCFA = .10). It should be noted 5-item measures have questionably lower convergent validity and severely under predict a few criteria.

**Conclusion**

The current study employed recent methodological advances in scale abbreviations using two meta-heuristics to develop personality assessments. Three primary findings emerged: First, the choice of algorithm carried the greatest consequence with very short scales. However, as the scales lengthened, all three algorithms converged in terms of dimensionality, reliability, and external validity. Second, different algorithms leveraged different psychometric criteria. A conventional SCFA produced short-forms which maximized homogeneity, whereas the GA better preserved convergent and external validity. ACO struck a balance by preserving acceptable levels of reliability and validity while also maximizing model fit. Therefore, effectiveness depends on the intended application of the abridged instrument. Finally, both automated algorithms compiled short versions that had improved factor structures, greater reliability, moresensitivity to personality differences, and similar validity in comparison to the longer scales. It should be noted that ACO was expected to outperform the GA, but the latter was far more time efficient (24 hours v. 5 minutes).

There are two major implications: First, results suggest the GA better preserves the properties of the initial test, whereas ACO will better mirror specifications of the user. In the current case, the HPI was initially designed for prediction and, as a result, was less focused on structural validity. The GA forms better captured the original HPI scores and, consequentially, may better preserve the empirical record of the instrument; at the same time, ACO tended to improve the unidimensionality and reliability of the forms, implying it may be better for individual decision-making and scale interpretation. Second, human test developers may turn to both ACO and GA to compile short versions that have, in comparison to longer forms, the same factor structure, reliability, sensitivity, and validity. This could facilitate implementation of organizational assessments by reducing participant burden without compromising scale validity.

**References**

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development*. Psychological assessment, 7*(3), 309–319.

Credé, M., Harms, P., Niehorster, S., & Gaye-Valentine, A. (2012). An evaluation of the consequences of using short measures of the Big Five personality traits. *Journal of Personality and Social Psychology*, *102*(4), 874–888.

Cronbach, L. J. (1954). Report on a psychometric mission to clinicia. *Psychometrika, 19*, 263– 270.

Haig, B. D. (2013). The philosophy of quantitative methods. In Little T. D. (Ed.), *The Oxford handbook of quantitative methods* (pp. 7-31). New York, NY: Oxford University Press.

Hogan, R., & Hogan, J. (1995). *Hogan Personality Inventory Manual* (2nd ed.). Tulsa, OK: Hogan Assessment Systems.

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.

Marcoulides, G.A., & Ing, M. (2012). Automated structural equation modeling strategies.

In R. Hoyle (Ed.). *Handbook of structural equation modeling*. New York, NY: Guilford Press.

Marsh, H. W., Morin, A. J. S., Parker, P. D., & Kaur, G. (2014). Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annual Review of Clinical Psychology, 10,* 85–110.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*, 1-36.

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological assessment, 8*(4), 350– 353.

Schroeders, U., Wilhelm, O., & Olaru, G. (2016). Meta-Heuristics in short scale construction: Ant colony optimization and genetic algorithm. *PloS one, 11*(11), e0167110.

Scrucca, L., and Sahdra, B. (2015). *GAabbreviate: Abbreviating Questionnaires (or*

*Other Measures) using Genetic Algorithms (Version 1.0)*: *R package*. Available

online at: http://CRAN.R-project.org/package=GAabbreviateYarkoni (2010)

**Table 1.** Model fit and McDonald’s ω of Original and Shortened Versions of the 7 HPI Scales.

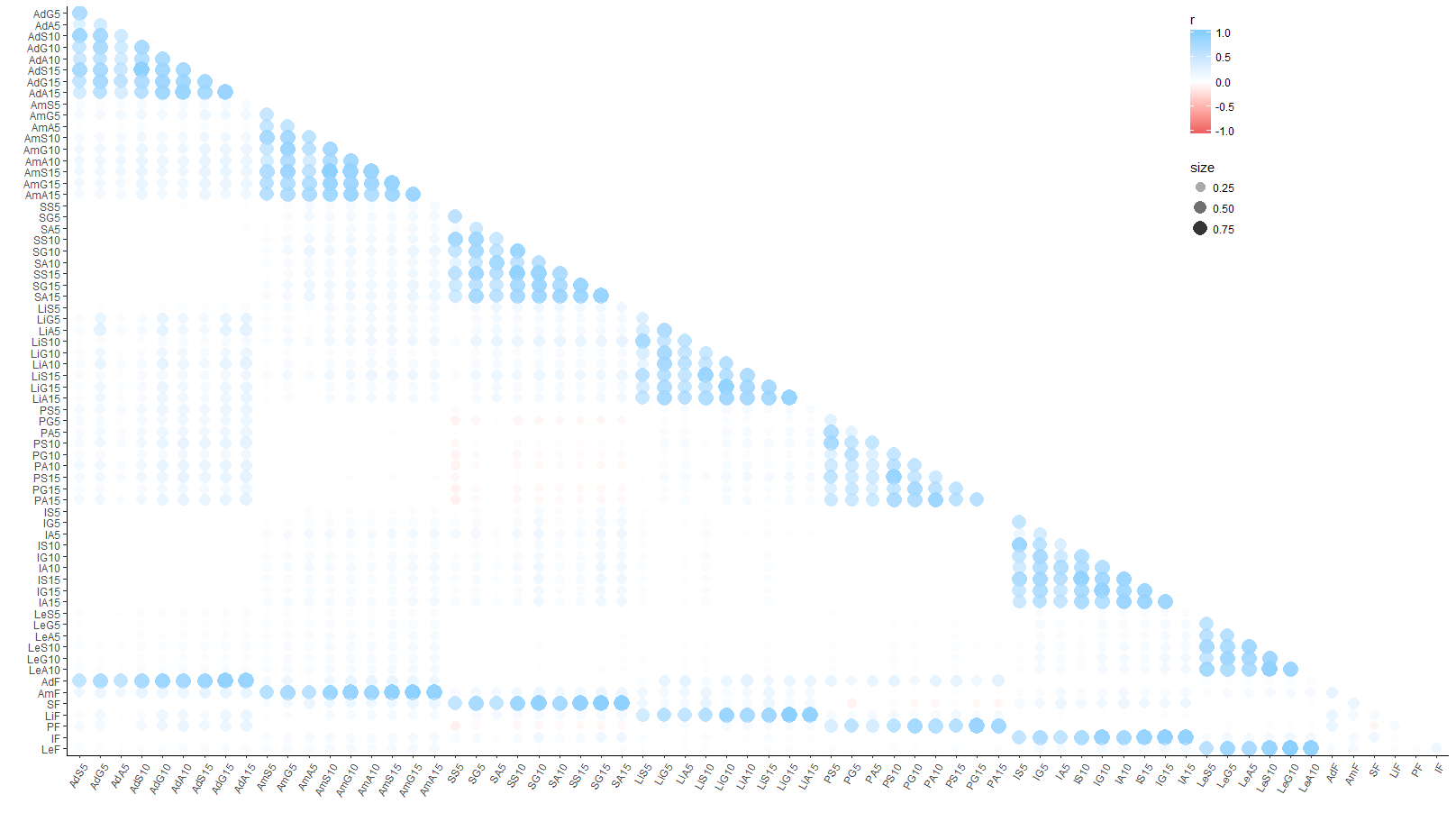
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Model** | ***df*** | **Χ2 WLSMV** | | | **CFI** | | | **RMSEA** | | | **McDonald’s ω** | | | | |
|  |  | **SC** | **GAA** | **ACO** | **SC** | **GAA** | **ACO** | **SC** | **GAA** | **ACO** | **SC** | **GAA** | | **ACO** | |
| Adj5 | 5 | 3270.36 | 49.1 | 26.45 | .910 | .988 | .997 | .226 | .026 | .018 | .82 | | .68 | | .75 |
| Adj10 | 35 | 8931.49 | 2153.54 | 565.64 | .821 | .861 | .964 | .141 | .069 | .034 | .87 | | .80 | | .80 |
| Adj15 | 90 | 12129.88 | 3087.29 | 1709.79 | .806 | .893 | .943 | .102 | .051 | .038 | .90 | | .84 | | .86 |
| Amb5 | 5 | 2765.14 | 325.2 | 20.76 | .854 | .943 | .999 | .208 | .071 | .016 | .88 | | .75 | | .82 |
| Amb10 | 35 | 7153.2 | 1739.22 | 565.64 | .790 | .889 | .964 | .126 | .062 | .034 | .89 | | .83 | | .84 |
| Amb15 | 90 | 10835.27 | 7999.68 | 2321.74 | .794 | .773 | .918 | .097 | .083 | .044 | .92 | | .88 | | .86 |
| Soc5 | 5 | 6364.84 | 264.72 | 40.67 | .902 | .932 | .997 | .316 | .064 | .024 | .90 | | .67 | | .82 |
| Soc10 | 35 | 17288.44 | 3425.99 | 770.07 | .787 | .831 | .962 | .197 | .087 | .041 | .88 | | .81 | | .80 |
| Soc15 | 90 | 29602.39 | 8371.35 | 3115.84 | .706 | .727 | .882 | .160 | .085 | .051 | .90 | | .81 | | .83 |
| Lik5 | 5 | 73.01 | 67.92 | 35.08 | .988 | .960 | .989 | .033 | .031 | .022 | .85 | | .61 | | .76 |
| Lik10 | 35 | 667.96 | 1343.69 | 304.22 | .943 | .776 | .945 | .038 | .054 | .025 | .88 | | .68 | | .80 |
| Lik15 | 90 | 2435.14 | 3329.01 | 1490.14 | .861 | .761 | .900 | .045 | .053 | .035 | .89 | | .78 | | .83 |
| Pru5 | 5 | 4125.35 | 78.05 | 44.45 | .814 | .967 | .998 | .254 | .034 | .025 | .84 | | .54 | | .65 |
| Pru10 | 35 | 6528.53 | 1585.85 | 868.44 | .761 | .755 | .937 | .121 | .059 | .043 | .79 | | .63 | | .72 |
| Pru15 | 90 | 11001.99 | 7189.66 | 2880.78 | .680 | .454 | .845 | .097 | .079 | .049 | .82 | | .65 | | .77 |
| Inq5 | 5 | 4592.43 | 151.89 | 45.15 | .783 | .959 | .987 | .268 | .048 | .025 | .83 | | .64 | | .64 |
| Inq10 | 35 | 8248.13 | 1841.23 | 1373.85 | .771 | .848 | .894 | .136 | .064 | .055 | .86 | | .73 | | .78 |
| Inq15 | 90 | 11932.14 | 6830.88 | 3108.69 | .754 | .760 | .861 | .102 | .077 | .051 | .86 | | .82 | | .81 |
| Lea5 | 5 | 2215.35 | 649.23 | 37.58 | .962 | .860 | .995 | .186 | .100 | .023 | .87 | | .67 | | .76 |
| Lea10 | 35 | 6828.33 | 4368.6 | 5586.59 | .902 | .754 | .919 | .123 | .099 | .111 | .86 | | .80 | | .85 |
| **Full Scale** |  |  | | |  | | |  | | |  | | | | |
| Adjustment | 629 | 33885.32 | | | .738 | | | .064 | | | .93 | | | | |
| Ambition | 377 | 20917.23 | | | .727 | | | .065 | | | .93 | | | | |
| Sociability | 252 | 38976.82 | | | .681 | | | .110 | | | .90 | | | | |
| Likeability | 209 | 5114.08  41427.6 | | | .774 | | | .043 | | | .87 | | | | |
| Prudence | 434 | .390 | | | .086 | | | .83 | | | | |
| Inquisitive | 275 | 27907.74 | | | .626 | | | .089 | | | .87 | | | | |
| Learning | 77 | 16128.18 | | | .796 | | | .128 | | | .86 | | | | |

For models, the first abbreviation reflects the parent scale: Adj = Adjustment, Amb = Ambition, Soc = Sociability, Lik = Likeability, Pru = Prudence, Inq = Inquisitive, Lea = Learning Approach. The second abbreviation the selection algorithm: SC = *Stepwise Confirmatory Factor Analysis,* GAA = *Genetic Algorithm Abbreviation*, ACO = *Ant Colony Optimization*. Finally, the number refers the item totals in the abridged versions.

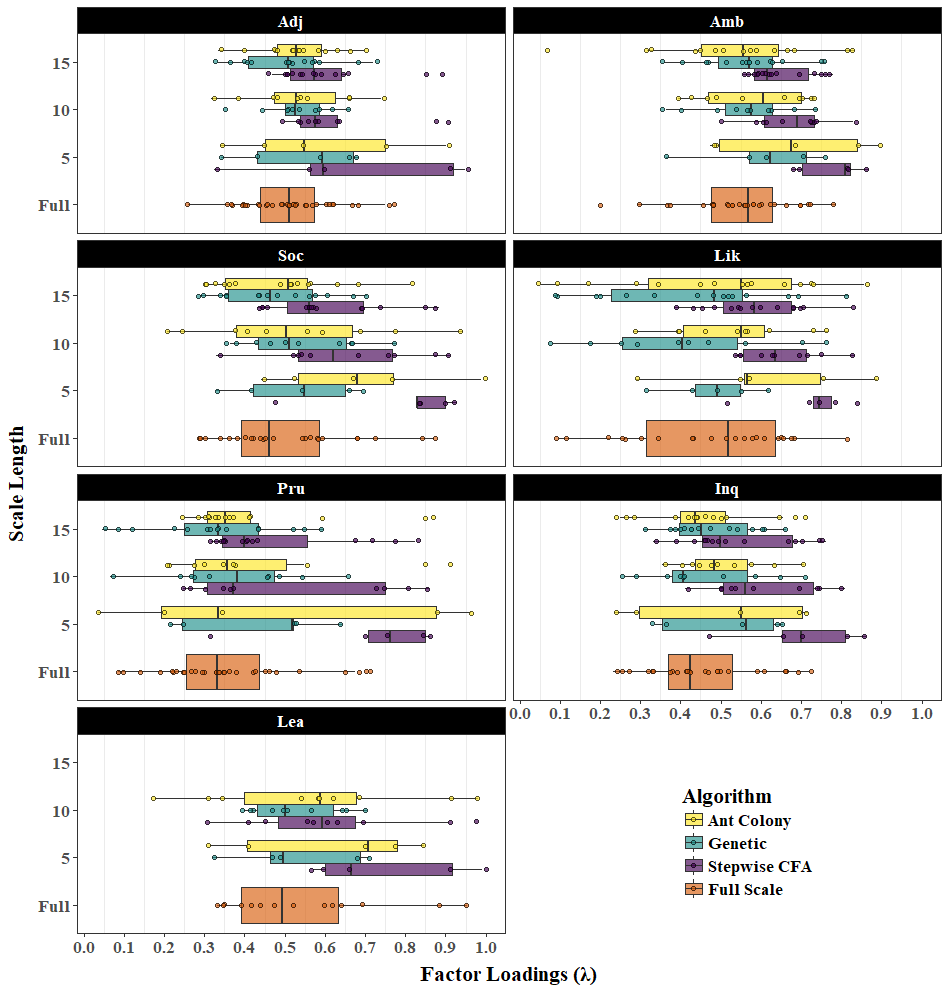
**Table 2.** Correlations of Abbreviated HPI scales with Full HPI scales and Differences in Correlations between the Abbreviated and Full HPI scales with Three Occupational scales (*n* = 12,761).

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Model** | **HPI Full** | | | **Reliability** | | | **Manager** | | | **Sales** | | |
|  | **SC** | **GAA** | **ACO** | **SC** | **GAA** | **ACO** | **SC** | **GAA** | **ACO** | **SC** | **GAA** | **ACO** |
| Adj5 | .72 | .81 | .71 | -.06 | -.06 | **-.36** | **-.14** | **-.12** | **-.21** | **-.11** | -.06 | -.05 |
| Adj10 | .84 | .90 | .86 | -.04 | -.01 | -.01 | -.06 | -.08 | -.02 | -.05 | -.03 | .00 |
| Adj15 | .89 | .94 | .91 | -.03 | -.03 | .00 | -.02 | -.04 | .00 | -.04 | -.01 | .02 |
| Amb5 | .77 | .83 | .76 | -.03 | -.02 | **-.18** | **-.12** | **-.17** | **-.20** | **-.24** | **-.11** | -.09 |
| Amb10 | .90 | .92 | .89 | -.02 | -.03 | .02 | **-.11** | -.09 | **-.17** | -.05 | -.03 | -.02 |
| Amb15 | .94 | .96 | .93 | .01 | -.02 | -.03 | -.09 | -.05 | -.03 | -.02 | .00 | -.03 |
| Soc5 | .70 | .83 | .72 | -.02 | .05 | **.18** | **-.16** | -.07 | -.03 | **-.31** | **-.17** | **-.26** |
| Soc10 | .89 | .93 | .86 | .02 | .06 | .04 | -.06 | .01 | -.05 | **-.12** | -.05 | **-.15** |
| Soc15 | .94 | .96 | .95 | .04 | .04 | .03 | -.05 | .01 | .00 | -.09 | -.03 | -.03 |
| Lik5 | .64 | .77 | .72 | **-.25** | -.03 | -.02 | -.02 | -.02 | .01 | .06 | -.02 | .01 |
| Lik10 | .78 | .90 | .84 | **-.21** | .02 | -.01 | .02 | -.05 | .00 | **.14** | -.07 | -.01 |
| Lik15 | .88 | .97 | .93 | **-.11** | .00 | -.02 | .05 | .00 | .00 | **.11** | .00 | .03 |
| Pru5 | .56 | .74 | .56 | **-.27** | -.09 | **-.34** | **-.12** | **-.17** | -.09 | **.14** | -.05 | **.14** |
| Pru10 | .74 | .85 | .78 | **-.17** | -.05 | .00 | -.09 | -.06 | -.01 | **.11** | .00 | .00 |
| Pru15 | .77 | .91 | .86 | **-.23** | -.04. | -.01 | -.06 | -.08 | -.03 | **.17** | -.01 | -.01 |
| Inq5 | .70 | .81 | .71 | -.01 | .01 | -.02 | **-.12** | -.06 | .00 | **-.12** | **-.10** | -.01 |
| Inq10 | .81 | .91 | .86 | -.02 | .01 | -.01 | -.09 | -.04 | .01 | -.08 | -.06 | -.01 |
| Inq15 | .91 | .95 | .93 | -.01 | .01 | .00 | -.01 | -.04 | .01 | .00 | -.04 | -.02 |
| Lea5 | .73 | .88 | .82 | .00 | -.03 | -.02 | -.03 | -.06 | .00 | **-.13** | -.07 | -.03 |
| Lea10 | .91 | .97 | .93 | .01 | -.01 | .01 | .01 | -.02 | .01 | -.02 | .00 | .00 |

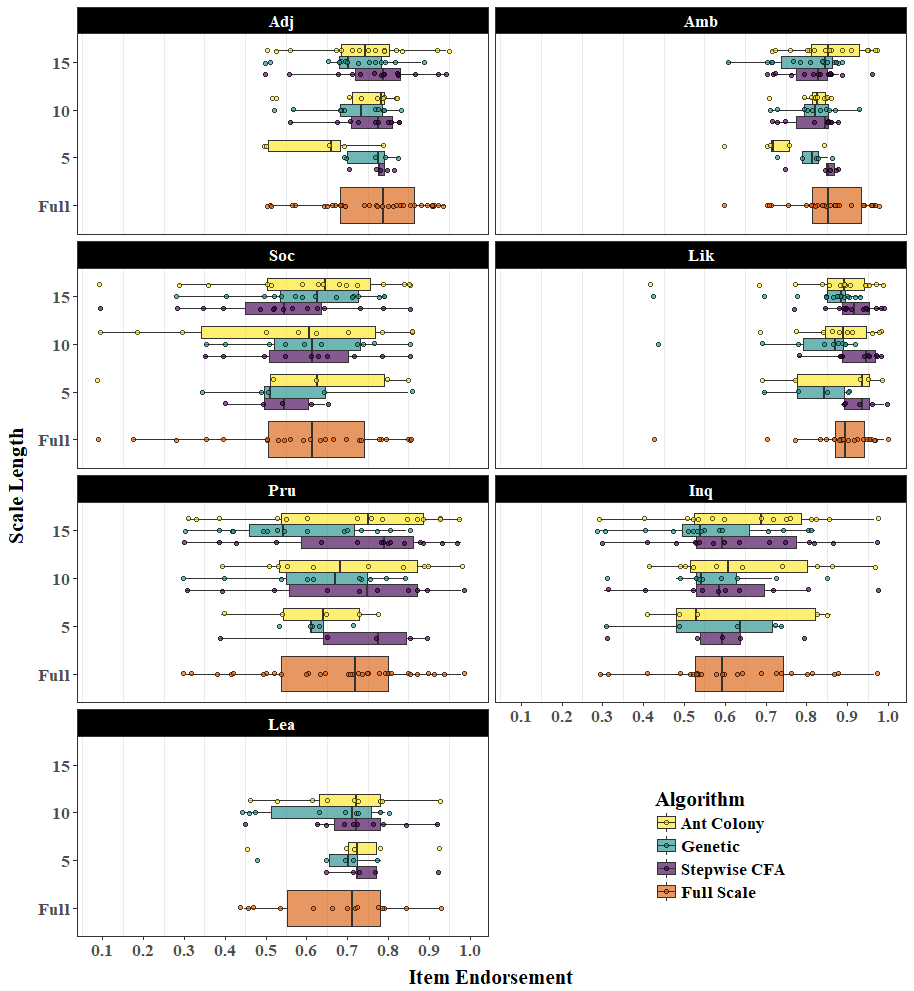
*Note.* Letters indicate item selection algorithm: SC = *Stepwise Confirmatory Factor Analysis;* GAA = *Genetic Algorithm;* ACO = *Ant Colony Optimization*; Adj = Adjustment; Amb = Ambition; Soc = Sociability; Lik = Likeability; Pru = Prudence; Inq = Inquisitive; Lea = Learning Approach. Scale abbreviations correspond to the HPI scale and number of items in the shortened form. Difference correlations = abbreviated HPI correlation – full HPI correlation. All discrepancies > | .10 | are bolded in black.



**Figure 1.** Correlogram for HPI Short and Full forms in the validation sample (N = 12,761).Variables abbreviated as follows: first 1-2 letters signify HPI scale, Ad = Adjustment, Am = Ambition, S = Sociable, L = Likeability, P = Prudence, I = Inquisitive, Le = Learning Approach; following the HPI abbreviation, the next capitalized letter signifies the algorithm for scaled reduction, S = Stepwise Confirmatory Factor Analysis, G = Genetic Abbreviation Algorithm, A = Ant Colony Optimization, F = Full Scale (original); finally, the number signifies the total number of items for the shortened scale, 5 = 5-item, 10 = 10-item, 15 = 15-item. A full correlation matrix with coefficients available upon request.



*Figure 2.* Distribution of Factor Loadings for Shortened and Full Scales. A boxplot cover the interquartile range; the solid line within a boxplot refers to the median. Numbers for scale length refer to the number of items in the abridged versions. Adj = Adjustment; Amb = Ambition; Soc = Sociability; Lik = Likeability; Pru = Prudence; Inq = Inquisitive; Lea = Learning Approach.



*Figure 3.* Distribution of Item Endorsement of Shortened Scales. A boxplot covers the interquartile range; the solid line within a boxplot refers to the median. Numbers for scale length refer to the number of items in the abridged versions. Adj = Adjustment; Amb = Ambition; Soc = Sociability; Lik = Likeability; Pru = Prudence; Inq = Inquisitive; Lea = Learning Approach.