Comparison of Automated Short Form Selection Strategies

Anthony Raborn[1] Walter Leite Katerina Marcoulides

Research and Evaluation Methodology Department

University of Florida

1: anthony.w.raborn@gmail.com

November 14, 2018

Applications of Psychometric Scales

Applied researchers are often faced with a dilemma, both with drawbacks:

A. Use a well-established but lengthy scale
   – Potentially longer administration time for less information

B. Use a few items from a scale
   – Potentially greater information but weaker validity evidence

In the literature, researchers attempt to use Option B with some effort spent on collecting validity evidence

Examples of Item Selection Methods for Short Forms

1. Hand-Selecting Items
   - Using theoretical or practical justifications per item (e.g., Noble, Jensen, Naylor, Bhullar, & Akeroyd, 2013)
   - Retaining one of many (qualitatively) redundant items (e.g., Dennis, 2003)
2. Statistical Criteria
   - Retaining items with high factor loadings or item correlations (e.g., Byrne & Pachana, 2011; Wester, Vogel, O'neil, & Danforth, 2012)
   - Selecting items that improve measures of reliability and/or dimensionality (e.g., Lim & Chapman, 2013; Veale, 2014)

Problem

Creating short forms with (1) good internal structure and (2) good predictive, convergent, and/or divergent validity is difficult by hand using *any* criteria.

One potential solution would be to use metaheuristic optimization algorithms (Dréo, Pétrowski, Siarry, & Taillard, 2006).

Goals of this Study

- Compare different automatic scale reduction strategies
  1. Model fit of final scales (better fit is better)
  2. Removal of specific problematic items (fewer problematic items is better)
  3. Reliability of final scales (higher reliability is better)
  4. Time to converge (faster is better)
- Determine which factors affect these comparisons
  1. Population model type (one factor, three factor)
  2. Severity of problematic items (none, minor, major)
  3. Strength of relationship to external criterion (none, moderate)

Previous Attempts

Some "common" algorithms in the literature:

1. "Maximize Main Loadings" (not investigated)
2. Ant Colony Optimization (ACO)
3. Tabu Search (TS)
4. Genetic Algorithm (GA)

An additional method investigated in this study:

5. Simulated Annealing (SA)

Factors Manipulated

1. The dimensionality of the full form
   - One Factor
   - Three Factor
2. Full-scale model misspecification
   - No misspecification
   - Minor misspecification (six items loading on a nuisance parameter with $\lambda = .3$)
   - Major misspecification (six items loading on a nuisance parameter with $\lambda = .6$)
3. Relationship to External Criterion Variable
   - No relationship
   - Moderate relationship ($\gamma = .6$)

One Factor Model

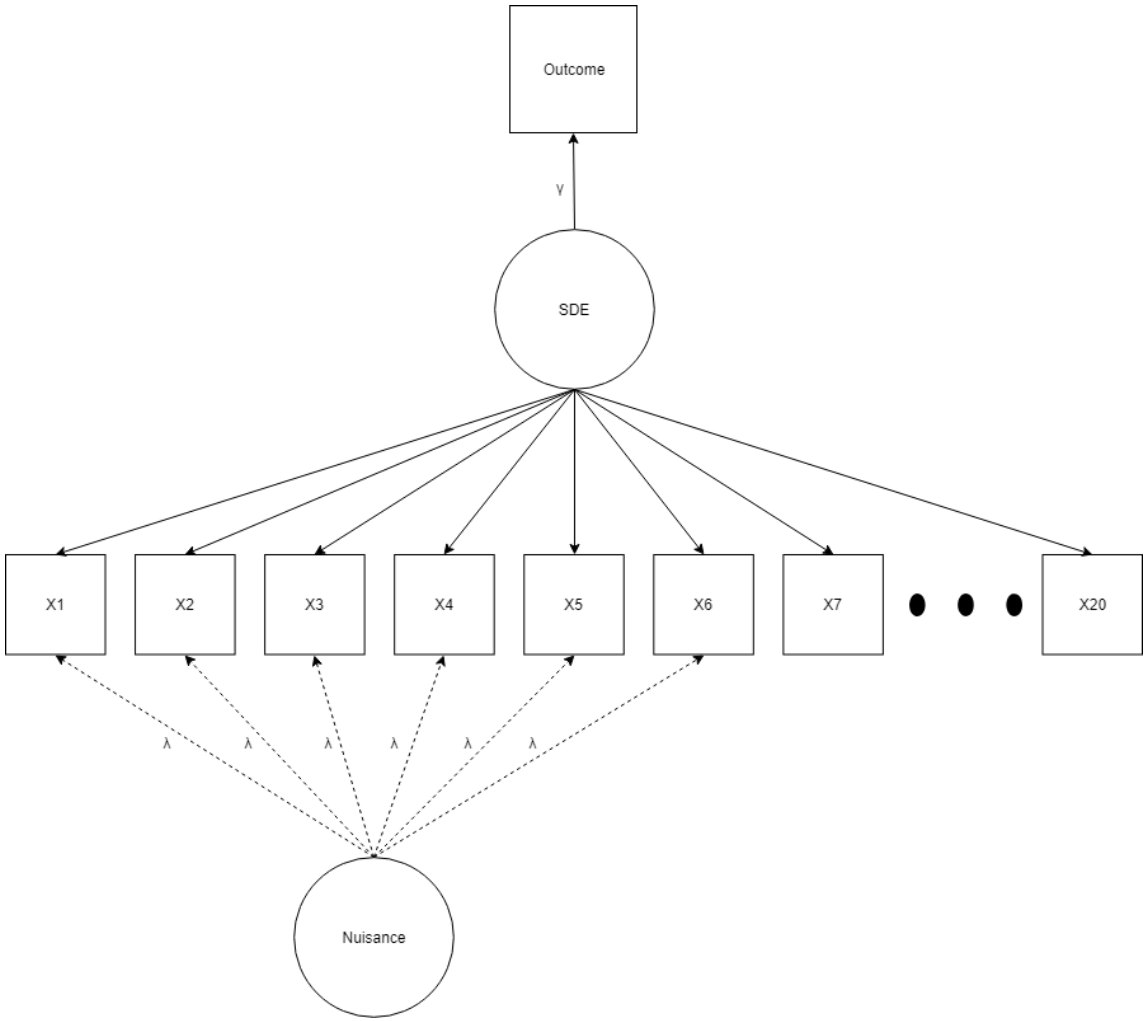

Figure 1: 20-item Self-Deceptive Enhancement Scale (Leite & Beretvas, 2005)

Three Factor Model

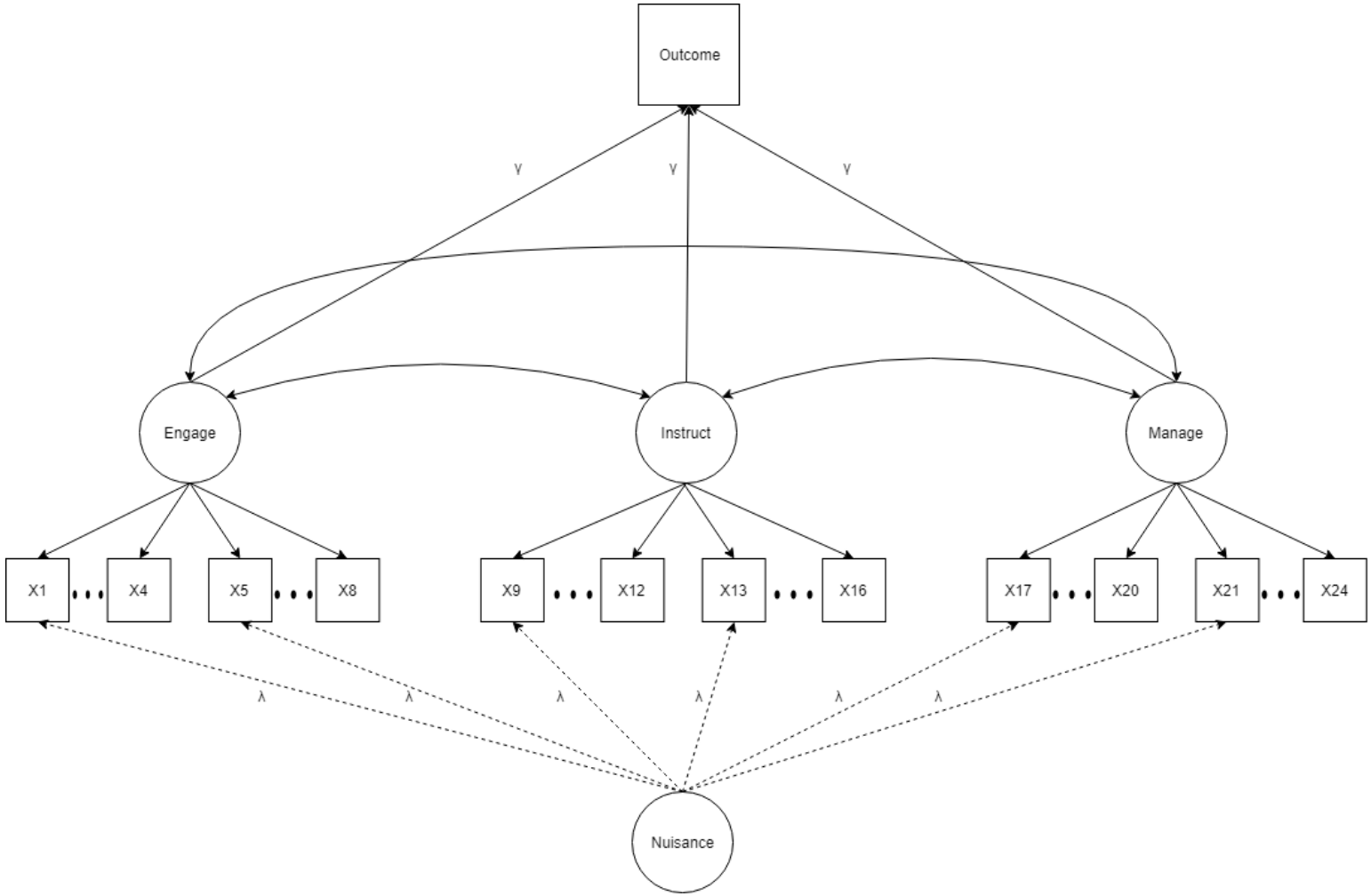

Figure 2: 24-item Teacher Efficacy Scale (Tschannen-Moran & Hoy, 2001)

Simulation

Program: R (R Core Team, 2018)

Packages:

1. `MASS` (Venables & Ripley, 2002) (data simulation)
2. `ShortForm` (Raborn & Leite, 2018) (ACO, SA, TS)
3. `GAabbreviate` (Sahdra et al., 2016) (GA; modified)

Sample Size: $n = 500$

Iterations: 100

Analysis of Results

1. CFI, TLI, RMSEA
2. Proportion of iterations including each problematic item (excluding no error condition)
3. Composite reliability of each factor:

$$CR_{factor} = \frac{(\Sigma_{i=1}^{I} Loading_i)^2}{(\Sigma_{i=1}^{I} Loading_i)^2 + \Sigma_{i=1}^{I}(Residual_i^2)}$$

4. Run time of algorithms

One Factor Model Fit: No External Variable

Error Condition	Method	CFI	TLI	RMSEA
None	ACO	**0.975**	**0.967**	**0.045**
	SA	**0.992**	**0.990**	**0.020**
	TS	**0.985**	**0.981**	**0.027**
	GA	**0.975**	**0.968**	**0.043**
Minor	ACO	**0.966**	**0.956**	0.052
	SA	**0.989**	**0.987**	**0.022**
	TS	**0.978**	**0.972**	**0.035**
	GA	**0.968**	**0.959**	**0.048**
Major	ACO	0.944	0.928	0.062
	SA	**0.983**	**0.979**	**0.028**
	TS	0.944	0.928	0.057
	GA	0.846	0.802	0.113

One Factor Model Fit: External Variable

Error Condition	External Relationship	Method	CFI	TLI	RMSEA
None					
	None	ACO	**0.975**	**0.967**	**0.045**
	None	SA	**0.992**	**0.990**	**0.020**
	None	TS	**0.985**	**0.981**	**0.027**
	None	GA	**0.975**	**0.968**	**0.043**
	Moderate	ACO	**0.979**	**0.973**	**0.040**
	Moderate	SA	**0.991**	**0.989**	**0.021**
	Moderate	TS	**0.985**	**0.981**	**0.027**
	Moderate	GA	**0.975**	**0.968**	**0.044**
Major					
	None	ACO	0.944	0.928	0.062
	None	SA	**0.983**	**0.979**	**0.028**
	None	TS	0.944	0.928	0.057
	None	GA	0.846	0.802	0.113
	Moderate	ACO	0.945	0.931	0.058
	Moderate	SA	**0.981**	**0.977**	**0.029**
	Moderate	TS	0.930	0.912	0.063
	Moderate	GA	0.858	0.822	0.107

One Factor Item Selection Proportions: No External Variable

Error Condition	Item	Factor Loading	ACO	SA	TS	GA
Minor	y3	0.580	0.84	0.39	0.48	0.66
	y5	0.534	0.81	0.36	0.43	0.35
	y4	0.448	0.70	0.59	0.53	0.61
	y2	0.408	0.66	0.36	0.37	0.41
	y6	0.393	0.51	0.38	0.39	0.93
	y1	0.382	0.34	0.39	0.48	0.48
Major	y3	0.580	0.49	0.12	0.46	0.61
	y5	0.534	0.35	0.32	0.48	0.44
	y4	0.448	0.33	0.17	0.48	0.70
	y2	0.408	0.19	0.19	0.37	0.48
	y6	0.393	0.21	0.21	0.36	0.96
	y1	0.382	0.21	0.21	0.48	0.44
Minor Error Average Proportion:			**0.643**	**0.412**	**0.447**	**0.573**
Major Error Average Proportion:			**0.297**	**0.203**	**0.438**	**0.605**

One Factor Item Selection Proportions: External Variable

Error Condition	External Condition	Item	Factor Loading	ACO	SA	TS	GA
Major	None	y3	0.580	0.78	0.48	0.46	0.82
	None	y5	0.534	0.75	0.39	0.47	0.98
	None	y4	0.448	0.65	0.45	0.50	0.21
	None	y2	0.408	0.34	0.34	0.41	0.29
	None	y6	0.393	0.29	0.55	0.55	0.29
	None	y1	0.382	0.16	0.48	0.51	0.18
Major	Moderate	y3	0.580	0.49	0.10	0.44	0.80
	Moderate	y5	0.534	0.17	0.13	0.42	0.98
	Moderate	y4	0.448	0.16	0.13	0.38	0.26
	Moderate	y2	0.408	0.12	0.14	0.47	0.43
	Moderate	y6	0.393	0.16	0.20	0.40	0.43
	Moderate	y1	0.382	0.15	0.24	0.53	0.46
	No External Average Proportion:			**0.495**	**0.448**	**0.483**	**0.462**
	Moderate External Average Proportion:			**0.208**	**0.157**	**0.440**	**0.560**

Three Factor Model Fit: No External Variable

Error Condition	Method	CFI	TLI	RMSEA
None	ACO	**0.980**	**0.974**	**0.042**
	SA	**0.992**	**0.990**	**0.023**
	TS	**0.989**	**0.986**	**0.027**
	GA	**0.979**	**0.973**	**0.042**
Minor	ACO	**0.972**	**0.964**	**0.050**
	SA	**0.990**	**0.987**	**0.026**
	TS	**0.984**	**0.979**	**0.035**
	GA	**0.970**	**0.961**	**0.050**
Major	ACO	**0.953**	0.939	0.062
	SA	**0.989**	**0.986**	**0.027**
	TS	**0.961**	**0.950**	0.053
	GA	0.909	0.882	0.089

Three Factor Model Fit: External Variable

Error Condition	External Relationship	Method	CFI	TLI	RMSEA
None					
	None	ACO	**0.980**	**0.974**	**0.042**
	None	SA	**0.992**	**0.990**	**0.023**
	None	TS	**0.989**	**0.986**	**0.027**
	None	GA	**0.979**	**0.973**	**0.042**
	Moderate	ACO	**0.977**	**0.970**	**0.047**
	Moderate	SA	**0.988**	**0.984**	**0.032**
	Moderate	TS	**0.983**	**0.978**	**0.037**
	Moderate	GA	**0.977**	**0.970**	**0.046**
Major					
	None	ACO	**0.953**	0.939	0.062
	None	SA	**0.989**	**0.986**	**0.027**
	None	TS	**0.961**	**0.950**	0.053
	None	GA	0.909	0.882	0.089
	Moderate	ACO	**0.964**	**0.953**	0.057
	Moderate	SA	**0.984**	**0.979**	**0.036**
	Moderate	TS	**0.953**	0.939	0.061
	Moderate	GA	0.907	0.880	0.091

Three Factor Item Selection Proportions: No External Variable

Error Condition	Item	Factor Loading	ACO	SA	TS	GA
Minor	y1	0.9	0.83	0.36	0.49	0.94
	y5	0.7	0.13	0.58	0.50	0.24
	y9	0.9	0.52	0.34	0.41	0.87
	y13	0.7	0.20	0.42	0.48	0.08
	y17	0.9	0.85	0.25	0.28	0.65
	y21	0.7	0.45	0.36	0.44	0.39
Major	y1	0.9	0.60	0.32	0.44	0.90
	y5	0.7	0.22	0.22	0.41	0.12
	y9	0.9	0.20	0.12	0.24	0.97
	y13	0.7	0.10	0.11	0.32	0.03
	y17	0.9	0.61	0.10	0.30	0.60
	y21	0.7	0.30	0.10	0.27	0.41
Minor Error Average Proportion:			**0.497**	**0.385**	**0.433**	**0.528**
Major Error Average Proportion:			**0.338**	**0.162**	**0.330**	**0.505**

Three Factor Item Selection Proportions: External Variable

Error Condition	External Condition	Item	Factor Loading	ACO	SA	TS	GA
Major	None	y1	0.9	0.60	0.32	0.44	0.90
	None	y5	0.7	0.22	0.22	0.41	0.12
	None	y9	0.9	0.20	0.12	0.24	0.97
	None	y13	0.7	0.10	0.11	0.32	0.03
	None	y17	0.9	0.61	0.10	0.30	0.60
	None	y21	0.7	0.30	0.10	0.27	0.41
Major	Moderate	y1	0.9	0.37	0.25	0.51	0.84
	Moderate	y5	0.7	0.15	0.12	0.37	0.16
	Moderate	y9	0.9	0.32	0.16	0.30	0.97
	Moderate	y13	0.7	0.06	0.16	0.20	0.03
	Moderate	y17	0.9	0.29	0.10	0.37	0.96
	Moderate	y21	0.7	0.04	0.09	0.37	0.04
	No External Average Proportion:			**0.338**	**0.162**	**0.330**	**0.505**
	Moderate External Average Proportion:			**0.205**	**0.147**	**0.353**	**0.500**

Best Performing Methods

1. Removal of specific problematic items: **SA**
2. Model fit of final scales: **SA**
3. Reliability of final scales: About equivalent (*ACO* somewhat higher)
4. Time to converge: About equivalent (*TS* one factor longer)

Overall: SA consistently had good[1] fit; ACO & TS consistently had at least adequate[2] fit; GA produced poor fit in the presence of major error

[1] CFI $>$ .95, TLI $>$ .95, RMSEA $<$ .05
[2] CFI $>$ .90, TLI $>$ .90, RMSEA $<$ .08

Factors Affecting Comparisons

1. Population model type
   - model fit: one factor $<$ three factor
   - Minimal effect on time to converge
2. Severity of problematic items
   - Decreased model fit, **SA** excluded
3. Strength of relationship to external criterion
   - Somewhat attenuates effect of error only for **ACO**

For the Future

Suggestions for Applied Researchers

A. Apply each algorithm to your sample–grab some coffee or tea while they each run!
B. Compare the resulting short forms against one another ("face validity" comparisons).
C. When possible, test against a second sample (cross-validation).

Future Research Questions

1. How well do the short forms created by each algorithm generalize to new samples?
2. How do additional manipulations (e.g., population models, types of errors) affect the algorithms?

Corresponding Author

anthony.w.raborn@gmail.com

References I

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.

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.

Dréo, J., Pétrowski, A., Siarry, P., & Taillard, E. (2006). *Metaheuristics for hard optimization: Methods and case studies*. Springer Science & Business Media.

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.

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.

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.

Raborn, A., & Leite, W. (2018). *ShortForm: Automatic short form creation*. Retrieved from https://github.com/AnthonyRaborn/ShortForm

R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

References II

Sahdra, K., B., Ciarrochi, J., Parker, P., . . . L. (2016). Using genetic algorithms in a large nationally representative american sample to abbreviate the Multidimensional Experiential Avoidance Questionnaire. *Frontiers in Psychology*, *7*(189), 1–14. Retrieved from http://www.frontiersin.org/quantitative_psychology_and_measurement/10.3389/fpsyg.2016.00189/abstract

Tschannen-Moran, M., & Hoy, A. W. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education*, *17*(7), 783–805.

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.

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with s* (Fourth). New York: Springer. Retrieved from http://www.stats.ox.ac.uk/pub/MASS4

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.