



The abbreviation of personality, or how to measure 200 personality scales with 200 items

Tal Yarkoni

Department of Psychology and Institute of Cognitive Science, University of Colorado at Boulder, UCB 345, Boulder, CO 80309, United States

ARTICLE INFO

Article history:
Available online 13 January 2010

Keywords:
Personality
Genetic algorithms
Abbreviation
Big Five
Measurement

ABSTRACT

Personality researchers have recently advocated the use of very short personality inventories in order to minimize administration time. However, few such inventories are currently available. Here I introduce an automated method that can be used to abbreviate virtually any personality inventory with minimal effort. After validating the method against existing measures in Studies 1 and 2, a new 181-item inventory is generated in Study 3 that accurately recaptures scores on eight different broadband inventories comprising 203 distinct scales. Collectively, the results validate a powerful new way to improve the efficiency of personality measurement in research settings.

© 2010 Elsevier Inc. All rights reserved.

1. Introduction

Filling out personality measures takes time. Unfortunately, time is a precious commodity in most psychological studies, and investigators who wish to comprehensively assess their participants' personalities must balance that wish against the need to allocate sufficient time to other components of a study. To reduce administration time, researchers often advocate the use of shortened versions of common personality measures (Ames, Rose, & Anderson, 2006; Francis, Brown, & Philipchak, 1992; Gosling, Rentfrow, & Swann, 2003; Robins, Hendin, & Trzesniewski, 2001; Saucier, 1994). Most prominently, the "Big Five" domains of personality are now routinely measured using 20-item, 10-item or even 5-item personality inventories in place of much longer measures (Donnellan, Oswald, Baird, & Lucas, 2006; Gosling et al., 2003; Langford, 2003; Rammstedt & John, 2007). Such inventories allow investigators to measure participants' personalities in a fraction of the time required to administer the original inventory, while recapturing scores in the original inventory relatively accurately.

Despite the potential benefits of using shorter measures, widespread development and use of abbreviated versions of existing measures is hindered by a number of limitations. First, development of an abbreviated measure can be a relatively laborious process. Most approaches to measure abbreviation emphasize the same type of conceptual and psychometric considerations commonly employed in development of entirely new measures (Smith, McCarthy, & Anderson, 2000; Stanton, Sinar, Balzer, & Smith, 2002), thus often necessitating considerable human input at various stages of development (e.g., multiple cycles of item selection,

analysis, and evaluation). The ability to automate much of the abbreviation process could potentially dramatically reduce the time investment required to generate short forms of existing measures.

Second, most short forms of existing measures are not guaranteed to achieve optimality, because their developers typically consider only a small fraction of possible alternate forms. For example, a common strategy for abbreviating multi-dimensional personality inventories is to select a subset of items from each scale that maximize item-total correlations while maintaining high internal consistency (e.g., Cox & Alexander, 1995; Goldberg et al., 2006; Lang & Stein, 2005; Troidahl & Powell, 1965); however, this approach can inadvertently restrict the breadth of the abbreviated measure by failing to sample from the full breadth of the original domain (Loevinger, 1954; Smith et al., 2000). Optimal abbreviation of a scale instead requires one to consider individual items not only in isolation, but also in combination with one another (Stanton, 2000). Ideally, optimality would be achieved by performing an exhaustive search through all possible permutations of items and retaining only the best combinations (e.g., Stanton et al., 2002); unfortunately, such an approach is often not computationally viable for longer measures, because of an inevitable combinatorial explosion (e.g., there are $3e + 25$ ways to choose a subset of 30 items from a larger pool of 100).

Third, the degree of length reduction possible for most multi-dimensional personality inventories is fundamentally limited in most studies by the adoption of two constraints: (a) the emphasis on selecting non-overlapping subsets of items for different scales (Budescu & Rodgers, 1981; Goldberg, 1972) and (b) the need to measure each construct with multiple items in order to ensure reliability and breadth of coverage (Nunnally, 1978). For example, for a 30-scale inventory such as the NEO-PI-R, even an abridged

E-mail address: tal.yarkoni@colorado.edu

inventory with only four items per scale (e.g., McCrae & Costa, 2007) will contain 120 items in total, a length that remains impractical in many testing situations. In order to achieve still greater reductions in measure length, one would presumably have to relax at least one of the two aforementioned constraints.

Fourth, researchers have focused largely on abbreviating low-dimensional personality inventories that are already relatively short (e.g., five or fewer dimensions, with not more than 100 items) and, with few exceptions (e.g., McCrae & Costa, 2007), have neglected longer broadband inventories such as the 240-item NEO-PI-R (Costa & McCrae, 1992) or 192-item HEXACO-PI (Lee & Ashton, 2004). The ability to abbreviate longer inventories could potentially enable much more efficient measurement of personality, as researchers would be able to obtain comprehensive assessments of personality in situations where time constraints currently preclude such an option.

Finally, because short-form personality inventories tend to be composed of items drawn entirely from their long-form counterparts, they are often protected by copyright and subject to the same terms of use as the original proprietary measures. Thus, researchers are often unable to create, modify or adapt short forms of measures that they do not themselves own, and may even be required to pay to use other researchers' measures. The ability to create abbreviated versions of existing inventories composed entirely of public-domain items would alleviate this limitation and facilitate the development and use of such measures.

The present article reports three empirical studies that collectively validate a novel approach to measure abbreviation capable of overcoming all of the above limitations. Rather than employing a conventional psychometric approach to measure development, the present studies make use of a genetic algorithm (GA)—a programmatic tool for solving high-dimensional problems (e.g., selecting a small number of variables from a much larger initial pool). Study 1 reports the automated development of a very short (10-item) Big Five inventory that accurately recaptures variance in a longer Big Five measure. Study 2 extends the same approach to the 240-item NEO-PI-R, demonstrating that a GA approach can be used to substantially shorten not only relatively short measures of personality but also broadband measures that assess many different facets of personality. Finally, Study 3 uses a GA to generate a single 181-item measure that can simultaneously recapture variance in eight different broadband personality inventories comprising 203 distinct scales. Collectively, Studies 1–3 demonstrate that the length of most personality inventories in common use can be reduced substantially with little loss in validity, to the point where multiple broadband inventories can potentially be replaced by a single relatively brief measure.

2. Genetic algorithm

2.1. Genetic algorithm-based item selection

In order to reduce a relatively long personality measure to a smaller subset of items, one must have a systematic way of determining how many, and which, items to retain. For longer measures, this may be difficult to do using existing approaches. For example, suppose that one wishes to abbreviate a 100-item inventory assessing 10 different scales to a much shorter inventory containing around 20 items. As discussed above, an exhaustive search approach is unlikely to identify the optimal solution in a reasonable amount of time due to the sheer number of possible permutations; conversely, conventional approaches based on careful consideration of psychometric desiderata are viable, but are likely to require considerable time and effort on the part of human researchers, and may produce sub-optimal results.

An alternative approach is to rely on a guided search algorithm that can identify near-optimal solutions while avoiding the problem of combinatorial explosion, by searching only sensible parts of the solution space. The present investigation relies on one particular class of such techniques known as *genetic algorithms* (GAs). Rather than relying on either brute force (searching every single possible combination) or sophisticated analytical approaches, GAs rely on simple evolutionary principles to produce high-quality solutions with few assumptions and relatively little computational labor. Because GAs may be unfamiliar to many personality psychologists, I provide a brief overview here. More detailed introductions can be found elsewhere (Davis, 1991; Mitchell, 1998; Whitley, 1994).

A GA has several essential components. First, it requires that the problem space be representable in a relatively compact form. Typically, this is a short string of characters (often referred to as *genes*), each of which represents a different parameter or variable of interest. These strings are usually referred to as *chromosomes*. In the present case, the N items of a given personality measure are represented as a chromosome consisting of N binary (0 or 1) genes that represent items, where each gene indicates whether the item is to be included (1) or excluded (0) from the resulting questionnaire. Thus, for example, the sequence 0 0 1 1 0 ... would indicate that, of the first five items on the original questionnaire, only items 3 and 4 are to be included in the final measure.

Given a suitable encoding of the problem space, one can then generate a starting *population* of chromosomes, typically in the range of 100–200 individuals. In the present studies, the initial population always consisted of 200 randomly generated chromosomes. This population provides the GA with an initial pool of genetic variability to operate with. However, variability alone is not enough to produce an adequate solution to most problems; the GA additionally requires some way to evaluate the quality, or *fitness*, of a given chromosome so as to select only those that show desirable properties. This requires the definition of a *fitness function* that takes an arbitrary chromosome as input, evaluates it against one or more criteria, and returns an overall index of its fitness. The specific function used in the present study is described in the next section.

Given an initial population and a defined fitness function, the GA then proceeds in a manner crudely analogous to evolutionary processes: in each generation, some proportion of the fittest chromosomes (often around 20%) are selected for breeding, and the rest are removed from the pool. The selected chromosomes are used to repopulate the entire pool of individuals in the next generation. To promote genetic diversity and ensure that the fitness of the population can continue to improve, mutation (spontaneous changes in a chromosome's genes) and recombination (two chromosomes exchanging genetic sequences) are introduced between generations with some specified frequency. The entire process is typically iterated for many (i.e., 100+) generations, at which point the fittest chromosome in the last generation is usually adopted as the solution to the problem. Note that GAs are not guaranteed to arrive at the single best solution (and will rarely do so); however, if the problem is not too complex, and the encoding and fitness functions are well defined, the final solution usually performs extremely well in practice—often better than solutions derived through much more thought- or computation-intensive means.

2.2. Fitness function

In the present studies, there were two important criteria for evaluating chromosomal fitness: (a) the number of items selected and (b) the amount of variance that the selected items explained in the original measure's various dimensions. The perfect measure would have as few items as possible while explaining as much of

the variance as possible in the original measure. Because these two constraints are in tension with one another, the central question is which of the two constraints to privilege. Should one aim for an extremely short measure that does a passable job at recapturing the original measure, or a longer one that almost perfectly recapture the original? The present GA was designed to accommodate either emphasis by varying a single parameter. The fitness function was defined as:

$$\text{Cost} = lk + \sum_{i=1}^s 1 - R_i^2$$

where l reflects a fixed item cost, k is the number of items retained by the GA, s is the number of scales in the inventory, and R_i^2 is the amount of variance in the i th scale that can be explained by a linear combination of individual item scores (see below). Note that the fitness function is defined in terms of cost, so that the goal of the GA is to find the solution that *minimizes* the fitness function. Thus, by varying l , one can place a greater or lesser emphasis on the brevity of the measure relative to its comprehensiveness. When l is high, the cost of each additional item outweighs the cost of a loss in explained variance, leading to a relatively brief measure. Conversely, when l is relatively low, the GA has little incentive to remove items, leading to a longer measure that maximizes explained variance.

One simple way to maximize the fit between the observed trait scores and the predicted trait scores (and hence, to minimize cost) would be to regress the observed trait scores on *all* of the items retained by the GA (e.g., given a 100-item inventory of five scales, use all 100 items to predict each of the five scale scores). However, this approach would produce scoring equations that are far too cumbersome to use in practice, involving dozens if not hundreds of items with non-unit coefficients (e.g., total score = $0.34 * \text{Item 1} + 0.71 * \text{Item 2} - 0.14 * \text{Item 3}$, etc.). To ensure that the GA produced measures with simple scoring equations, an alternative approach was adopted. For each scale, only the five items that correlated most strongly with total scores of the target scale were used in scoring (in Study 1, only the top two items were used, in order to produce a very brief measure). The sum of these five item scores was then used to predict the actual (i.e., observed) personality scores, and the squared correlation was taken as the R^2 value used in the above function. This approach ensured that only items with non-negligible correlations with the targeted scale were used to predict scores, while keeping the resulting scoring equations manageable. Note that no prohibition against item overlap was imposed; an individual item could be used to score more than one scale, thereby enabling the GA to minimize redundancy between items.

The GA simulations were implemented in the statistical package *R* (Team, 2008) using the free *genalg* library (Willighagen, 2005). In all studies, the population consisted of 200 chromosomes. The number of generations over which the simulation was run varied across studies, since the search was terminated once the cost function stabilized (i.e., when running additional iterations did not further minimize cost, or did so almost imperceptibly). The parameter l , reflecting the item cost, varied across studies as described below.

3. Study 1

The goal of Study 1 was to validate a GA approach to inventory abbreviation by determining whether a standard Big Five measure—in this case, the 44-item Big Five Inventory (BFI; John, Donahue, & Kentle, 1991)—could be substantially shortened with minimal loss of measurement accuracy. In contrast to previous studies that developed short personality measures using standard psychometric procedures (e.g., Gosling et al., 2003), items were se-

lected for inclusion in the present measure in an atheoretical manner and with no consideration for standard psychometric criteria. Instead, a genetic algorithm was used to produce a very brief measure that empirically optimized the balance between the length of the measure and its ability to recapture the dimensions of the original BFI.

3.1. Method

3.1.1. Participants and measure

Data for Study 1 were provided by Gosling and colleagues (Gosling et al., 2003), and are from a previous study validating the Ten Item Personality Inventory (TIPI), a very brief Big Five inventory. One hundred and fourteen participants filled out the 44-item Big Five Inventory (BFI) on two occasions two weeks apart. For full details, see Gosling et al. (2003).

3.1.2. GA settings

In Study 1, l , which represents the item cost, was set to 0.05—a value that tended to produce measures in the targeted range of 10 items.

3.2. Results

The GA selected 10 of the 44 items in the BFI for inclusion in the abbreviated measure.¹ Table 1 displays the correlations between Big Five scores as assessed by the 10-item versus 44-item measures. Convergent correlations between the full-scale BFI and abbreviated version ranged between .83 and .93 for the five dimensions (mean = .87); conversely, none of the off-diagonal correlations exceeded .27 (mean absolute $r = .11$). Table 1 also displays estimates of internal consistency (Cronbach's alpha) and test-retest reliability for the 10-item and 44-item versions of the measure. Note that Cronbach's alpha tends to underestimate reliability for short and/or heterogeneous measures (Cronbach, 1951; Osburn, 2000); thus, test-retest coefficients provide a more accurate estimate of reliability in this case. Importantly, the latter were only slightly lower for the abbreviated measure (mean = .75) than for the full-scale BFI (mean = .83).

3.3. Discussion

The results of Study 1 provided clear evidence that it is possible to programmatically generate a very short Big Five inventory that measures personality almost as reliably as a measure that is more than four times as long. It is important to note that the measure produced in Study 1 was not intended to compete with or replace existing very short Big Five inventories, and there is no reason to believe it offers any advantages over existing measures such as the TIPI (Gosling et al., 2003). Rather, Study 1 served as a proof of concept that it is possible to develop high-quality abbreviated measures using an automated rather than rational approach. The major benefit of an automated approach is that it could, in principle, be used to produce abbreviated versions of other existing personality measures relatively effortlessly, whereas a rational approach requires considerable time and effort. In particular, as the complexity of the target measure grows (e.g., broadband inventories containing a dozen scales or more), a rational approach could quickly become impractical, whereas an automated approach

¹ Because the goal of Study 1 was strictly to validate the present methodology, and not to produce another measure that would compete with existing inventories (Gosling et al., 2003; Langford, 2003; Rammstedt & John, 2007), the item list and scoring key for the 10-item measure are not provided in the present article. However, they are available from the author upon request.

Table 1
Relationship between the 10-item GA-based Big Five measure and the 44-item Big Five inventory (BFI).

GA-based measure	Correlation with full-scale BFI scores					Cronbach's alpha		Test-retest correlation	
	E	A	C	N	O	GA	BFI	GA	BFI
Extraversion	0.93	0	0	−0.02	−0.11	0.77	0.87	0.85	0.85
Agreeableness	0.03	0.86	0.23	−0.21	−0.15	0.47	0.79	0.72	0.84
Conscientiousness	0	0.23	0.83	−0.17	0.08	0.39	0.77	0.64	0.81
Neuroticism	0.04	−0.25	−0.17	0.9	−0.08	0.64	0.83	0.77	0.82
Openness	−0.04	−0.11	−0.03	−0.27	0.83	0.57	0.81	0.79	0.84

should remain equally viable. Study 2 was designed to empirically test the latter prediction.

4. Study 2

The goal of Study 2 was to produce an abbreviated version of the 240-item NEO-PI-R, a widely used broadband personality inventory that assesses 30 distinct 'facets' of personality rather than just five broad domains. A similar approach to Study 1 was used to demonstrate that a much shorter (in this case, 56-item), programmatically generated, version of the measure could reliably recapture the original 30 facet scales.

Because the NEO PI-R is a proprietary measure, Study 2 used a large set of public-domain items to create the abbreviated NEO PI-R analog. That is, rather than selecting a subset of NEO PI-R items that most closely approximated total scores on the 30 scales, an independent set of items were substituted for the NEO PI-R items. This approach ensured that researchers could freely use the resulting measure (cf. Goldberg, 1999; Goldberg et al., 2006), while demonstrating the power of the GA-based approach to automatically recapture scores on a given inventory using a relatively arbitrary pool of items.

4.1. Method

4.1.1. Personality data

Studies 2 and 3 used data from the Eugene–Springfield Community Sample (ESCS; Goldberg et al., 2006). Data were provided by Goldberg and colleagues at the Oregon Research Institute. Full details are reported in other publications (Goldberg et al., 2006; Grucza & Goldberg, 2007). Briefly, over a span of 15 years, over a thousand participants in the Eugene–Springfield area in Oregon have filled out many of the most widely used proprietary personality inventories (e.g., NEO-PI-R, HEXACO-PI, JPI-R, etc.), as well as over 2000 individual public-domain items that constitute the International Personality Item Pool (IPIP). Study 2 used data for the subset of ESCS participants ($N = 857$) who had completed the NEO PI-R.

4.1.2. GA settings

Because the large number of available IPIP items made the GA-based analysis impractical,² the pool of items was first reduced to a manageable set of 500 items. For each of the 30 facet scales, the 10 items that correlated most strongly with the total scale score were included; then, the remaining items were selected by ordering all items based on the number of different scales with which they correlated and selecting those items that correlated significantly ($p < .05$) with the greatest number of different scales. This approach ensured that the pool of 500 items contained items that were both strongly correlated with specific traits and diffusely cor-

related with many traits. The GA was then run on the 500-item set of IPIP items with I , the item cost, set to 0.05.

Because test–retest data were not available for the ESCS sample, an alternative cross-validation approach was used in Studies 2 and 3. To ensure that the measures produced by the GA gave unbiased results and did not capitalize on chance, half of the available sample was used to train the GA, and the other half was used for testing purposes. That is, the scoring equations generated in the training sample were used to predict scores in the testing sample from item-level data, and the GA-predicted scores were then correlated with the "real" scores (i.e., the actual scores on the original personality measures such as the NEO-PI-R). Unless explicitly noted, all results are reported only for the testing sample, and hence are unbiased. Additionally, all R^2 values reported in regression analyses in Studies 2 and 3 are adjusted for the number of predictors in order to correct for shrinkage.

4.2. Results

The NEO-PI-R analog produced by the GA contained only 56 items—a reduction of 77% in length from the original.³ To evaluate the quality of this measure, two sets of analyses were conducted. First, I directly computed the similarity between the GA-based and original versions of the NEO-PI-R; second, I assessed the degree to which the two versions exhibited similar patterns of correlation with external variables.

4.2.1. Similarity of original and GA-based measures

To provide a simple metric of the similarity between the two measures, the convergent correlation between the two versions of the measure was computed for each of the 30 facets and five domains. At the facet level, the mean convergent correlation was .57 (Table 2). Although this value may seem low in comparison to the results of Study 1, it is important to note that narrowband personality measures such as the NEO-PI-R facets tend to be somewhat less reliable than broad domains such as the Big Five (Costa & McCrae, 1988). Moreover, in Study 1 the items selected by the GA were a strict subset of the items on the BFI, and were administered in the same session, whereas the IPIP items in Study 2 did not overlap with the NEO-PI-R items, and were administered on multiple occasions up to 13 years after the NEO-PI-R. Thus, the present estimates were likely to be substantially attenuated due to any limitations in the long-term stability of the facet scores.⁴

At the domain level, convergent correlations were higher, with a mean of .7. These correlations were in line with previous test–retest correlations reported for Big Five domains (Costa & McCrae, 1988; McCrae, 2001; for a more general review, see Roberts & DelVecchio, 2000), suggesting that there was relatively little loss

² Because most IPIP items were only weakly related to the NEO-PI-R facets, including these items in the search space would have forced the GA to search through parts of the space that were unlikely to produce good results, needlessly lengthening the search time.

³ A scoring key for this measure is not provided here, as the 108-item version of the NEO PI-R produced in Study 3 (Appendix A) was deemed superior. However, the key is available from the author upon request.

⁴ In fact, the present values were only slightly lower than previously reported test–retest correlations of between .63 and .83 for the NEO-PI-R facets at shorter intervals of 3 or 6 years (Costa & McCrae, 1988).

Table 2
Relationship between the NEO PI-R and GA-based analog.

	Cronbach's alpha		Convergent correlation	CV correlation	CV correlation for BRI	CV correlation for peer BFI
	Original	GA				
<i>Neuroticism</i>			0.78	0.98	0.98	.98
Anxiety	0.83	0.67	0.67	0.97	0.92	.97
Angry hostility	0.8	0.55	0.64	0.95	0.92	.96
Depression	0.85	0.71	0.68	0.96	0.97	.98
Self-consciousness	0.74	0.68	0.72	0.97	0.94	.99
Impulsiveness	0.72	0.27	0.5	0.96	0.85	.96
Vulnerability	0.79	0.65	0.68	0.96	0.93	.94
<i>Extraversion</i>			0.72	0.97	0.87	.97
Warmth	0.8	0.57	0.64	0.94	0.91	.94
Gregariousness	0.8	0.62	0.68	0.91	0.86	.98
Assertiveness	0.8	0.4	0.62	0.93	0.89	.86
Activity	0.72	0.51	0.57	0.89	0.88	.92
Excitement-seeking	0.64	0.43	0.52	0.77	0.53	.74
Positive emotions	0.81	0.49	0.61	0.94	0.9	.93
<i>Openness</i>			0.67	0.92	0.94	.99
Fantasy	0.82	0.29	0.43	0.86	0.62	.94
Aesthetics	0.84	0.49	0.59	0.82	0.94	.96
Feelings	0.75	0.49	0.51	0.89	0.79	.99
Actions	0.64	0.44	0.5	0.93	0.87	.94
Ideas	0.82	0.59	0.63	0.95	0.88	.93
Values	0.78	0.25	0.54	0.9	0.92	.87
<i>Agreeableness</i>			0.64	0.91	0.96	.99
Trust	0.84	0.58	0.61	0.95	0.93	.99
Straightforwardness	0.74	0.4	0.44	0.9	0.89	.92
Altruism	0.72	0.51	0.51	0.86	0.92	.93
Compliance	0.73	0.47	0.5	0.91	0.87	.94
Modesty	0.75	0.37	0.5	0.84	0.85	.94
Tender-mindedness	0.61	0.49	0.48	0.87	0.89	.96
<i>Conscientiousness</i>			0.71	0.97	0.93	.99
Competence	0.7	0.57	0.5	0.94	0.88	.90
Order	0.74	0.56	0.66	0.91	0.89	1.00
Dutifulness	0.67	0.55	0.4	0.91	0.85	.99
Achievement striving	0.67	0.59	0.56	0.92	0.87	1.00
Self-discipline	0.8	0.61	0.64	0.97	0.9	.97
Deliberation	0.7	0.44	0.45	0.88	0.57	.92
Mean for 30 facets	0.76	0.51	0.57	0.92	0.89	.94
Mean for five domains	–	–	0.7	0.95	0.94	.98

of precision associated with the use of a restricted set of IPIP items in place of the full NEO-PI-R. Importantly, domain scores were produced by summing the six facet scales within each domain; the GA was not given the original domain scores as input. Thus, the fact that aggregating over facet scores increased convergent validity provided independent evidence that the GA was recapturing reliable variance in each facet and was not simply capitalizing on chance.

In addition to correlating each of the 30 facet and five domain scores across the two versions of the measure, I assessed the extent to which inter-facet correlations in the GA-based measure recaptured the pattern observed in the original measure. For each NEO-PI-R facet, all correlations with other facets were normalized using Fisher's r -to- z transformation, and the column–vector correlations between the two measures were computed (cf. Funder & Sneed, 1993; Gosling et al., 2003). The mean column–vector correlation across the 30 facets was .92, indicating a very high degree of similarity between the correlation matrices.

4.2.2. Correlations with external criteria

To assess the degree to which the GA-based and original versions of the NEO-PI-R showed similar correlations with external variables, I conducted analyses comparing the similarity of the correlations between the two versions of the NEO-PI-R and an independent set of criterion variables. Specifically, I related the 30 facets of the NEO-PI-R to (a) the Behavioral Report Inventory (BRI), a set of 60 clusters of behavioral acts defined by Goldberg

and colleagues based on frequency ratings of 400 activity descriptions (e.g., borrowed money, chewed gum, donated blood, played chess, etc.; Grucza & Goldberg, 2007) and (b) peer ratings of personality provided by 2–3 knowledgeable informants using the Big Five inventory (John et al., 1991).⁵ The BRI clusters reflected highly specific real-world behaviors (e.g., drug use, travel, sleeping, and housekeeping) that should show relatively selective associations with narrowly defined personality traits such as the NEO PI-R facets; the peer-rated BFI scores, which did not rely on self-report, were used to ensure that the convergence of the original and GA-based measures held across different methods. Taken together, these variables provided a strong test of the extent to which the GA-based version of the NEO-PI-R maintained both the sensitivity and the specificity of the original measure.

For the BRI analysis, an initial column–vector correlation analysis revealed that the mean column–vector correlation between the two versions of the measure for the 30 facets was .89. That is, the pattern of correlations between the GA-based NEO PI-R analog scores and the 60 BRI clusters was very similar to the pattern of correlations between the original NEO-PI-R scores and the BRI clusters. Moreover, when each BRI cluster was simultaneously regressed on all 30 facets, the mean amount of explained variance

⁵ For each participant, ratings were provided by 1–3 peers, and these peers varied in the nature of their relationships to the target. For simplicity, and to increase reliability, I averaged over all available peer ratings within each subject prior to analysis.

(i.e., multiple R^2) across the 60 clusters was 15.4% for the original NEO-PI-R and 11.7% for the GA-based version. Thus, the GA-based version managed to explain three-quarters as much BRI variance as the original NEO-PI-R using only one-quarter as many items. Furthermore, the correlation between the two measures' ability to explain variance in the BRI was .80 across the 60 clusters, indicating that BRI clusters that could be better predicted by the NEO-PI-R were also better predicted by the GA-based measure.

When the same analyses were performed on the five dimensions of the peer-rated BFI scores, an even greater degree of convergence was observed (Table 2). The mean column–vector correlation between the NEO-PI-R and GA analog was .92, and when each of the BFI dimensions was regressed on all 30 facets, the GA analog explained a mean of 21.5% of the variance in peer-rated BFI scores, versus a mean of 25.9% for the NEO-PI-R.

4.2.3. Explicit manipulation of the brevity–fidelity tradeoff

Importantly, the reduction in predictive power observed for the GA-based measure relative to the NEO-PI-R in some analyses (e.g., a somewhat lower multiple R^2 when predicting BRI clusters and peer-rated BFI scores) was not intrinsic to the GA itself, and reflected the choice of item cost (i.e., the parameter I) stipulated in the fitness function. Because the goal was to generate a highly abbreviated measure, the item cost was initially set relatively high ($I = 0.05$) in order to emphasize a reduction in measure length. However, by assigning a lower value to I , one could easily produce longer measures that captured more of the variance in the original NEO-PI-R. For example, when I was set to 0.03 (i.e., a greater emphasis on fidelity relative to brevity), the resulting measure contained 79 items, and was more similar to the NEO-PI-R than the 56-item version, with a mean convergent correlation of .59 between facets of the two versions (versus .57 for the 56-item version), and a mean column–vector correlation of .92 with the BRI clusters (versus .89). When I was set to 0.01, the resulting 103-item measure had a mean convergent correlation of .61 and a mean column–vector correlation of .93. Moreover, the latter measure explained an average of 14.5% of the variance in the 60 BRI clusters and 23.4% of the variance in peer-rated BFI scores—only slight decrements from the 15.3% and 25.9% of the original NEO-PI-R. Thus, by varying the item cost parameter, one could easily produce somewhat longer versions of the NEO-PI-R that were better able to recapture the variance of the original while still remaining considerably shorter than the full 240-item NEO-PI-R or even its existing 120-item short form (McCrae & Costa, 2007).

4.3. Discussion

Study 2 demonstrated that most of the variance in the 30 facets and five domains of the 240-item NEO-PI-R could be successfully recaptured using just 56 items. For most practical purposes, the GA-based NEO-PI-R analog appeared to be interchangeable with the much longer NEO-PI-R. Moreover, the degree of convergence with the NEO-PI-R could be increased by arbitrarily varying the item cost in order to produce a somewhat longer measure. Thus, Study 2 suggested that researchers could optimize the fidelity–brevity tradeoff on a case-by-case basis to meet the constraints of individual studies, ensuring that participants' time can be used as efficiently as possible.

5. Study 3

The goal of Study 3 was to extend the approach used in Study 2 in two ways. First, Study 3 sought to abbreviate inventories other than the NEO-PI-R in order to demonstrate the generality of the GA-based approach. Second, Study 3 sought to increase the

efficiency of the abbreviation process by simultaneously recapturing multiple inventories using a single, relatively short, personality measure. To achieve these aims, I set a target of recapturing approximately 200 different personality scales using approximately 200 IPIP items. Inspection of the personality measures available in the ESCS sample suggested a combination of eight relatively well-known personality inventories that collectively included 203 different personality scales: the 30-scale NEO-PI-R (Costa & McCrae, 1992), 24-scale HEXACO-PI (Lee & Ashton, 2004), 31-scale Temperament and Character Inventory (TCI; Cloninger, 1994), 30-scale California Personality Inventory (CPI; Gough, 2000), 44-scale Hogan Personality Inventory (HPI; Hogan & Hogan, 1995), 15-scale Jackson Personality Inventory (JPI-R; Jackson, 1994), 11-scale Multi-dimensional Personality Questionnaire (MPQ; Tellegen & Waller, 1994), and 18-scale Six-Factor Personality Questionnaire (6-FPQ; Jackson, Paunonen, & Tremblay, 2000).⁶ Although there was substantial content overlap between these eight measures, many also included scales that their creators have argued are largely overlooked by other measures (e.g., the Honesty–Humility scales of the HEXACO-PI, or the self-transcendence scales of the TCI). Thus, the challenge in Study 3 was to generate a measure that not only captured the commonalities between measures, but also accurately preserved the specific components of each measure, using a limited number of items.

5.1. Method

The approach used in Study 3 was identical to that used in Study 2, with three exceptions. First, the sample contained only those ESCS participants who had filled out at least 90% of the 203 scales ($N = 545$)⁷; second, in order to maximize the breadth of coverage of the personality space, the pool of items included in the GA analysis contained 1000 items rather than 500 items. Third, the item cost was set to 0.03 instead of 0.05, following pilot testing indicating that this level produced measures of approximately 200 items in length.

5.2. Results

The abbreviated measure generated by the GA contained 181 items, a 91% reduction from the 209 items that would be required to administer all eight measures individually.⁸ Henceforth, this measure is referred to as the Analog to Multiple Broadband Inventories (AMBI). Items included in the AMBI are provided in Appendix A, and scoring equations for the 203 scales are provided in Appendix B. The convergent validity of the AMBI was evaluated using the same metrics as in Study 2.

5.2.1. Correlation between original and GA-based measures

Table 3 reports the mean convergent correlation for each of the eight constituent inventories, averaged across all scales on that inventory (convergent correlations for individual scales are reported in Appendix B). The obtained values ranged between .53 and .67, and were comparable to test–retest coef-

⁶ Some of the inventories administered to the ESCS sample were subsequently revised by their authors; thus, the number of scales available for some inventories diverges from the most recent published versions (e.g., the most widely used version of the TCI contains 25 subscales, reflecting the deletion of several of the facets included in the present version).

⁷ Missing scores were handled via pair-wise deletion.

⁸ The GA initially produced a 184-item measure; however, subsequent inspection revealed that three items were included twice each (the duplications occurred because the IPIP item pool contained several items with multiple unique IDs). These three items were therefore removed and scoring equations for all scales were recomputed using the remaining 181 items. All results are reported for the 181-item measure.

Table 3
Relation between AMBI and original personality inventories.

Measure	Scales	Items (orig.)	Items (GA)	Convergent corr.		CV corr.		CV corr. with BRI		CV corr. with peer-rated BFI	
				Scales	Domains	Scales	Domains	Scales	Domains	Scales	Domains
NEO-PI-R	30	240	108	0.63	0.8	0.92	0.98	0.89	0.96	0.94	0.98
HEXACO-PI	24	192	88	0.67	0.8	0.87	0.97	0.9	0.96	0.93	0.97
TCI	31	247	100	0.61	0.74	0.87	0.96	0.86	0.94	0.88	0.95
HPI	44	194	123	0.53	0.74	0.83	0.95	0.84	0.93	0.87	0.98
6-FPQ	18	108	69	0.58	0.69	0.8	0.94	0.83	0.87	0.89	0.93
CPI	30	462	58	0.55	–	0.84	–	0.85	–	0.88	–
JPI-R	15	300	66	0.63	–	0.83	–	0.89	–	0.94	–
MPQ	11	276	47	0.63	–	0.85	–	0.9	–	0.91	–

CV = column–vector; BRI = behavioral report inventory.

ficients reported for many of the same inventories in previous studies (e.g., Cloninger, Przybeck, Svrakic, & Wetzel, 1994; Gough & Bradley, 1996; Hansenne, Delhez, & Cloninger, 2005; Helson & Moane, 1987; Schuerger, Zarrella, & Hotz, 1989; Usala & Hertzog, 1991).

Interestingly, there was heterogeneity in convergent correlations both within and between measures. Some scales within a measure had lower correlations than others (Appendix B), and correlations were generally lower for the 6-FPQ, CPI, and HPI (mean r s = .53 to .58) than for the other measures (all r s > .61). Importantly, however, this heterogeneity appeared to reflect differences in the reliabilities of the original scales rather than limitations of the GA approach, as evidenced by a high correlation ($r = .69$) between the 203 convergent correlation coefficients and the internal consistency coefficients reported for the 203 scales on the IPIP website (reproduced in Appendix B). That is, those scales that showed lower convergent correlations tended to have relatively low internal consistency, suggesting that they were relatively unreliable to begin with.

Because 5 of the 8 inventories (all but the JPI-R, CPI and MPQ) were structured hierarchically, it was possible to test the correspondence between GA-based and original inventories not only at the level of individual scales but also at the level of higher-order domains. For all five measures, correlations between the two versions were higher for the higher-order domains (mean = .75) than for the individual scales (mean = .61; Table 3). As in Study 2, the higher-order domain scores were not explicitly included in the GA's fitness function, and the GA's ability to predict these domain scores accurately therefore derived from its capacity to reliably predict the corresponding lower-order scales.

To assess the similarity between the inter-scale correlation matrices of the original and AMBI measures, the mean column–vector correlation for each inventory was computed (Table 3). A high degree of similarity (>.8) was observed for all measures. The similarity of the correlation matrices was also readily apparent visually, as illustrated for the AMBI version of the NEO PI-R in Fig. 1.⁹ Moreover, absolute differences in correlations between the two original and GA-based measures were relatively small; Fig. 2 displays histograms quantifying changes in NEO-PI-R inter-facet correlations as a function of differences in samples (i.e., training versus validation halves of the ESCS sample; left panel), measures (i.e., original NEO-PI-R versus AMBI; center panel), or both (right panel). The figure demonstrates that the magnitude of differences in correlations between the original and AMBI versions was comparable to that observed between two sep-

arate NEO-PI-R samples,¹⁰ and that there was no evidence of any systematic inflation or deflation of correlations in the GA-based version relative to the original (i.e., the distribution of differences was symmetric and centered around 0).

5.2.2. Convergent correlations with external criteria

The same analyses as in Study 2 were used to assess the degree of convergent correlation between the original and AMBI inventories when predicting two sets of external variables (the BRI clusters and peer-rated BFI scores). Table 3 summarizes the results for each of the eight inventories. Overall, the AMBI scales showed a close resemblance to the original scales. column–vector correlations reflecting the similarity of correlation matrices between each version of the inventories were high for both the BRI clusters and peer-rated BFI scores (all > .83; Table 3; see also Supplementary Fig. S1 online). Finally, the constituent scales of each AMBI inventory explained approximately as much variance as the original inventories in both the BRI clusters (mean variance explained for the original eight inventories = 14.8%, mean for AMBI inventories = 13.9%) and peer-rated BFI scores (mean for original inventories = 24.9%, mean for AMBI inventories = 25%). The notable exception was the CPI, for which the original measure substantially outperformed the AMBI analog when predicting the BRI clusters (original $R^2 = .18$, AMBI $R^2 = .12$).

5.2.3. Multicollinearity of TCI self-transcendence subscales

For 200 of the 203 scales included in the AMBI inventory, the GA generated unique scoring equations as expected (i.e., each scale was made up of a unique combination of items). However, for three of the self-transcendence subscales on Cloninger's TCI (ST3, ST4, and ST5), the GA produced the same scoring equation for all three scales. Initially, this was flagged as a potential problem with the present methodology; however, subsequent inspection indicated that intercorrelations between these three subscales were in fact extremely high in the original TCI data (0.79, 0.82, and 0.9), suggesting that they measured essentially the same construct. Moreover, although the version of the TCI administered to ESCS participants contained five distinct self-transcendence subscales, in a subsequent revision (the TCI-R), the inventory's authors have retained only three subscales, removing ST4 and ST5 because of their redundancy with ST3 (Cloninger, 1994). Thus, the apparent multicollinearity of the GA-based ST subscales was taken to reflect

⁹ Due to space limitations, only the figure for the NEO PI-R is presented here; however, corresponding figures for the other inventories are available upon request from the author.

¹⁰ Note that this comparison is biased in favor of the NEO-PI-R, because participants filled out all NEO-PI-R items in a single session, whereas items on the AMBI were administered across multiple occasions spanning several years. Systematic differences in session-related variables (e.g., differences in mood) would be expected to contribute to inter-facet correlation patterns on the NEO-PI-R, but would have no such effect on the AMBI.

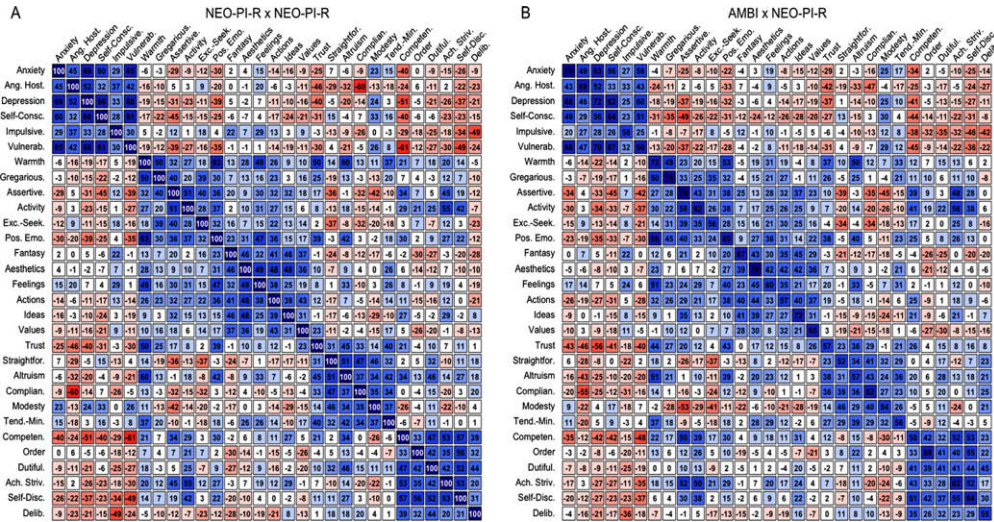


Fig. 1. Convergent correlations between AMBI version of the NEO PI-R and the proprietary version. Left: inter-facet correlation matrix for the 30 facets of the original NEO PI-R. Right: correlation matrix between the 30 AMBI facets (on the y-axis) and the 30 original facets (on the x-axis). Positive correlations are in light colors with black text (blue in the online version); negative correlations are in dark colors with white text (red in the online version). In the online version of the figure, darker colors indicate stronger correlations. Decimal points are omitted.

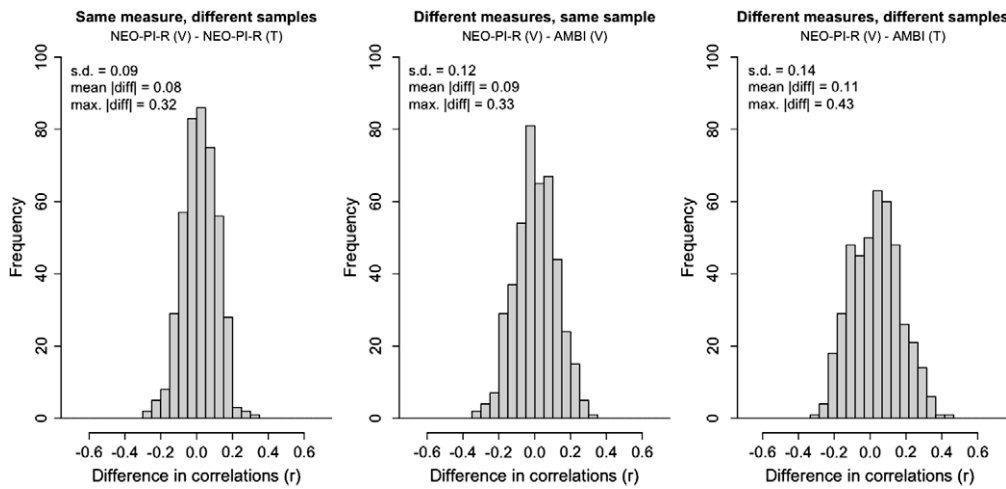


Fig. 2. Histograms of differences in inter-facet correlations between the NEO-PI-R and AMBI measures in the validation (V) and/or training (T) samples. Standard deviation and mean and maximum absolute differences in correlations are displayed for each distribution.

limitations of the original measure, and no further steps were taken.

5.2.4. Comparison with other automated approaches

Although the GA-based approach to scale abbreviation required minimal human intervention and ran relatively quickly as implemented,¹¹ it was important to ensure that it outperformed other even simpler approaches that could also be readily automated. I therefore contrasted the GA-based results with those produced by two alternative approaches. A first approach consisted of a brute-force search through a large set of randomly generated item subsets. This type of random sampling approach is the basis of a number of strategies for generating alternate forms (Armstrong, Jones, & Wu, 1992; Gibson & Weiner, 1998; Oswald, Friede, Schmitt, Kim, & Ramsay, 2005). Ten thousand samples of 200 items

were randomly drawn from the pool of 1000 IPIP items used to generate the AMBI, and the same cost function used for the GA in Studies 1–3 was used to evaluate the quality of the result (i.e., balancing the reduction in measure length against the inevitable drop in explained variance). Fig. 3 displays the distribution of costs produced using this approach. For comparison, the actual cost of the AMBI is also plotted. The cost of the actual AMBI solution was 8.06 standard deviations better than the mean random solution and 4.62 standard deviations better than the single best randomly-generated solution, suggesting that there was virtually no chance of a brute-force search outperforming the GA in any reasonable period of time.

A second automated approach selected items solely based on the strength of their item-total correlations with the target scales (e.g., Goldberg et al., 2006). For each of the 203 scales, the five most strongly correlated items were selected for inclusion in the final measure. This approach did in fact produce a measure that outperformed the AMBI; however, the gains were extremely modest (mean cross-validated convergent correlation = .61, versus .6 for the AMBI), and the slight increment in explained variance was achieved at the

¹¹ The AMBI took a few hours to generate on a 1.83 GHz Core Duo MacBook Pro laptop running an “off-the-shelf” genetic algorithm package (*genalg*), with virtually no human intervention required. The measures generated in Studies 1 and 2 each took only a few minutes to generate.

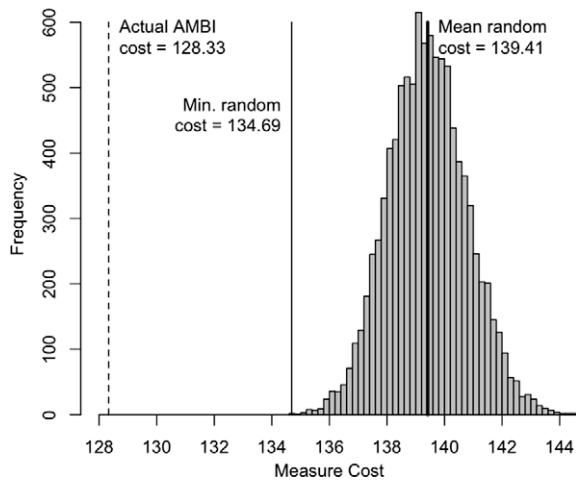


Fig. 3. Histogram showing the distribution of costs for 10,000 randomly generated 200-item measures. The actual AMBI cost is also plotted for reference. All costs are for the (unbiased) validation half of the ESCS sample.

cost of an unreasonably large increase in measure length, to 531 items. To compare measures of approximately equal lengths, I repeated the analysis with only the top two items selected for each scale. The resulting measure now contained only 264 items (still 80 more than the AMBI); however, its performance was also substantially poorer than the AMBI's (mean = .54 versus .6).

In sum, the GA clearly outperformed two alternative approaches that have been previously used to abbreviate personality measures. More sophisticated approaches capable of producing better results might be developed, of course; however, there is reason to believe that any further gains would be modest at best. As Fig. 4 illustrates, the GA's performance gains decayed exponentially across generations and asymptoted relatively quickly, suggesting that there is likely to be a relatively hard limit to any further cost reductions.

5.3. Discussion

Study 3 demonstrated that it was possible to reliably recapture most of the variance in multiple broadband inventories using a much smaller set of public-domain items. On average, each inven-

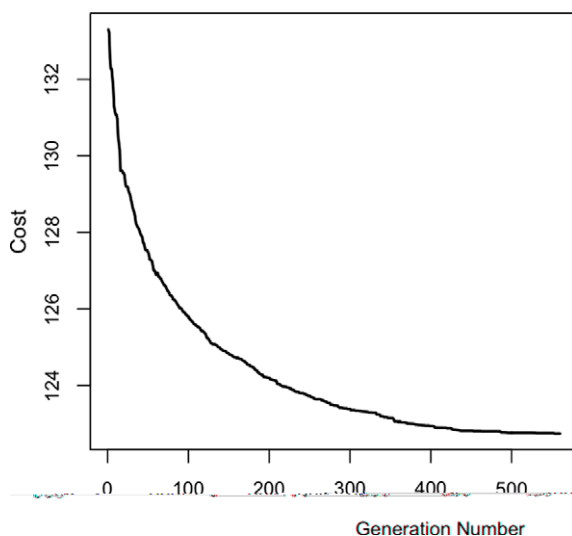


Fig. 4. Evolution of AMBI fitness across generations. Each point represents the cost associated with the single best measure in each generation of the training sample.

tory was measured using fewer than half the number of items used in the original inventory (Table 3); importantly, because items were allowed to load on multiple scales, the resulting AMBI measure contained only 181 unique items. Extensive validation analyses demonstrated that the use of the AMBI in place of the original inventories had minimal impact on results. In general, the AMBI scales correlated strongly with the corresponding original scales, exhibited very similar patterns of correlation with other variables, and explained approximately the same amount of variance in external variables as the original scales.

Interestingly, there was heterogeneity in the degree to which the AMBI inventories resembled the originals. Whereas some inventories (e.g., the NEO PI-R, HEXACO-PI, and TCI analogs) converged very closely with the original measures, others (e.g., the CPI and HPI) showed a weaker correspondence. As noted above, much of this heterogeneity appeared to reflect differences in the reliabilities of the original scales rather than limitations of the present methodology; for example, the HPI, which had the lowest convergent correlations with the AMBI analog (mean of 44 scales = .53; Table 3), also had the lowest reported internal consistency coefficients (mean = .61; Appendix B). Nonetheless, researchers may wish to exercise caution when using the AMBI analogs of the HPI and CPI—at least for scales that have relatively low convergent correlations (Appendix B). Fortunately, the same caution does not appear to be warranted for the AMBI analogs of inventories such as the NEO PI-R and HEXACO-PI, which Study 3 suggested could be substituted for the original inventories with very little impact on results.

6. General discussion

The present investigation validated the use of an algorithmic approach to personality measure abbreviation that is capable of substantially reducing a measure's length while retaining most of its predictive power. The chief implication of these findings is a practical one: researchers can use GA-based abbreviated measures to dramatically reduce the amount of time spent administering questionnaires while simultaneously broadening the number of personality dimensions assessed. In particular, the AMBI measure developed in Study 3 (Appendix A) contains only 181 items, yet accurately measures scores on eight different broadband inventories comprising 2019 items. Moreover, the measure is extremely flexible, in that researchers can select subsets of AMBI items that substitute for specific inventories (e.g., the 108 items used to produce NEO PI-R scores, the 88 items used to produce HEXACO-PI scores, or the 140-item intersection of the two measures). Widespread use of this measure, or of other measures developed using a similar approach, should greatly enhance the efficiency of personality data collection while ensuring that results from different studies remain commensurable.

More generally, the results validated a novel method for generating highly abbreviated versions of single or multiple existing measures. Because the GA-based approach is largely automated, it requires much less time and effort than more conventional approaches that involve rational deliberation. Moreover, as Study 3 demonstrated, a GA-based approach remains viable in cases where the complexity of the desired measure makes rational development difficult if not impossible. To facilitate the use and development of GA-based personality measures in future studies, I have developed a website (<http://www.shortermeasures.com>) that provides abbreviated versions of a number of widely used personality questionnaires. All items are drawn from the public domain IPIP pool (Goldberg et al., 2006), and automatic scoring routines for each measure are provided in both SPSS and R formats. Provisional R code for generating new GA-based measures is also provided.

In addition to their practical implications, the present findings also have two important implications for scale development and psychometric theory, as several features of the present approach appear to contradict conventional psychometric wisdom. First, the GA-based approach deliberately violated a common psychometric recommendation against using individual items more than once during scoring. The typical justification for this recommendation is that allowing items on different scales to overlap can spuriously inflate (or sometimes, deflate) the correlation between those scales due to shared error (Budescu & Rodgers, 1981; Goldberg, 1972; Kaplan & Saccuzzo, 1982). Although this concern is a valid one, it is important to recognize that one cannot entirely eliminate correlated error terms between multiple scales simply by using non-overlapping items. Many sources of correlated error require only the presence of content overlap between items—for example, replacing an item like “I often feel blue” with two separate items that have similar content (e.g., “I experience frequent bad moods” and “I am sad much of the time”) will not prevent a person in a transient bad mood from endorsing both items. Conversely, the introduction of new items carries certain costs; for example, in an attempt to avoid re-using the same item twice, researchers may re-word an item in such a way as to inadvertently introduce unrelated unique variance to a scale, spuriously decreasing the correlation between different scales. Ultimately, it is an empirical question as to whether item overlap presents a problem in any given instance. Empirically, Studies 2 and 3 found no evidence of meaningful inflation of inter-scale correlations in the GA-based measures relative to the original measures (such inflation would have manifested in low column–vector correlations in Tables 2 and 3, a wide distribution in Fig. 2, and would have been visually apparent in Figs. 1 and Supplementary Fig. S1). When one weighs the negligible changes observed in inter-scale correlation patterns against the increased brevity and flexibility of the present measures (e.g., the 91% reduction in measure length for the AMBI), it is difficult to argue that the costs of allowing item overlap outweigh the benefits.

Second, one might argue that the measures produced by the GA appear to be unreliable, because Cronbach's alpha, an estimate of internal consistency, was in the conventionally inadequate range (e.g., below .7; Smith et al., 2000) for most AMBI scales (mean = 0.59). This concern is empirically refuted inasmuch as the convergent validity of the resulting scales was clearly very high (i.e., GA-based measures showed virtually identical patterns of correlation with other variables as the original measures, and the vast majority of convergent correlations were higher than coefficient alpha), and validity necessarily implies reliability. What the low coefficient alpha values do underscore, however, is the underappreciated point that internal consistency estimates largely reflect the amount of redundancy present in a measure, and consequently tend to underestimate reliability for short or heterogeneous measures (Cronbach, 1951; Osburn, 2000). It is entirely possible to have a measure with very high reliability and validity that has very low internal consistency due to the presence of highly heterogeneous items (Cattell & Tsujioka, 1964; McClelland, 1980). Indeed, many researchers have argued that this is the ideal state of affairs, because it maximizes the efficiency of data collection while ensuring that the domain under investigation is richly sampled (Boyle, 1991; Cattell & Tsujioka, 1964; Kline, 2000; McDonald, 1981; Stanton et al., 2002). Thus, far from presenting a problem, low internal consistency coefficients in the present studies represent the inevitable end-state of successful measure abbreviation. High internal consistency estimates would imply that the GA had failed to do its intended job of eliminating redundant variance.

Of course, the GA approach is not without its limitations. First, in cases where fidelity to existing measures is of the utmost importance, researchers should not use measures such as the AMBI,

which necessarily entail some divergence from the original criterion scores, however small. For example, in cases where researchers wish to compare the distribution of personality scores in a sample with some other population's established NEO-PI-R norms, GA-based measures would obviously produce incommensurable results, due to differences in item selection and scaling. Fortunately, there are relatively few instances in which interpretation of absolute personality scores is desirable or warranted; indeed, some critics have objected to the very notion of a personality “norm” (Goldberg et al., 2006). The vast majority of personality analyses concern the covariation between personality scores and other variables, and in this respect, the present findings demonstrate that GA-based measures such as the AMBI can be reasonably substituted for the corresponding proprietary measures.

Second, because the present GA approach is wholly atheoretical, it should be used with great caution—or not at all—when developing entirely new measures or refining existing ones. The GA approach is appropriate if one's sole goal is to reduce the length of existing measures, as in the present study; however, it is important to recognize that any resulting measure will simply be a shorter facsimile of its long-form counterpart, and will inherit any conceptual or psychometric problems that affect the original measure. Development or refinement of personality measures should take into account theoretical considerations that a simple algorithmic approach cannot. A more sophisticated automated approach might eventually be developed that is capable of taking such considerations into account, but the present implementation lacks any such sophistication.

Finally, it is important to note that the GA approach is only one of many possible ways to automate the abbreviation of personality measures. As demonstrated in Study 3, the GA approach appears to outperform two simpler approaches that have been used in past research (i.e., selecting the top N items for each criterion scale, or selecting the best of a very large set of randomly-generated measures). However, there may well be other existing approaches the author is unfamiliar with that are superior to the GA approach in terms of either quality of results or ease of implementation. And even if such alternatives do not currently exist, they may well be developed in future. Although the speed with which the GA's cost function asymptoted in Study 3 (see Fig. 4) seems to suggest that any further benefits are likely to be incremental, no claim is made here that the GA approach represents the optimal method for automated measure abbreviation, but only that it is a readily available, easily implemented, and highly effective one. Should better alternatives emerge, researchers would do well to use those alternatives instead.

In sum, the present investigation validates a novel approach to the abbreviation of personality measures while introducing a new measure that can be used as a freely available substitute for many of the most widely used proprietary personality inventories. Of course, GA-based measures such as the AMBI are not perfect facsimiles of the inventories they are intended to recapture. In cases where fidelity to the original inventories is essential, researchers should continue to use the original inventories; and in cases where the goal is development or refinement of a measure rather than length reduction, researchers should continue to use existing psychometric techniques. However, in the many cases where the need for brevity outweighs the need for fidelity, and where researchers have no prior commitment to a specific form of a personality inventory, measures such as the AMBI provide a significant advance in the pursuit of efficient personality measurement.

Acknowledgments

This research was partially supported by NIH Grant F32NR012081. The author wishes to thank Sam Gosling, Lewis Goldberg, and Maureen Barckley, who provided the data used in Studies

1–3, and Nicholas Holtzman, who provided valuable feedback on an earlier version of the manuscript.

Appendix A

The 181 items on the AMBI. Alphanumeric labels in parentheses indicate the ID of each item in the IPIP database.

No.	IPIP ID	Description
1	(D79)	Rarely worry
2	(X145)	Often eat too much
3	(D70)	Usually like to spend my free time with people
4	(H334)	Take charge
5	(H58)	Am always busy
6	(X63)	Radiate joy
7	(X45)	Do not like poetry
8	(X12)	Distrust people
9	(A141)	Tell other people what they want to hear so that they will do what I want them to do
10	(E124)	Anticipate the needs of others
11	(X61)	Believe that I am better than others
12	(X259)	Sympathize with the homeless
13	(Q104)	Keep things tidy
14	(V98)	Am a highly disciplined person
15	(E46)	Jump into things without thinking
16	(A108)	Admire a really clever scam
17	(M58)	Like to own things that impress people
18	(Q238)	Try to be with someone else when I'm feeling badly
19	(E136)	Feel others' emotions
20	(R11)	Talk a lot
21	(S22)	Try to avoid speaking in public
22	(C4)	Am usually active and full of energy
23	(Q59)	Am patient with people who annoy me
24	(Q210)	Am easily annoyed
25	(Q204)	Push myself very hard to succeed
26	(X86)	Do not like art
27	(V330)	Love to hear about other countries and cultures
28	(R35)	Am considered to be kind of eccentric
29	(P420)	Worry about what people think of me
30	(V91)	Am very shy in social situations
31	(V211)	Don't have much energy
32	(H579)	Don't think that laws apply to me
33	(E92)	Have frequent mood swings
34	(X217)	Hold a grudge
35	(E122)	Like to act on a whim
36	(A47)	Have felt the presence of another person when he or she was not really there
37	(D20)	Seem to derive less enjoyment from interacting with people than others do
38	(P393)	Love to be the center of attention
39	(P378)	Don't let others cut in front of me in line
40	(V69)	Insist that others do things my way
41	(V223)	Am a firm believer in thinking things through
42	(X43)	Dislike changes
43	(V291)	Read a large variety of books
44	(X50)	Love excitement
45	(Q116)	Would rather spend money than save it
46	(E35)	Enjoy being reckless
47	(V187)	Have never cared much what others thought of me
48	(D40)	Usually get right to work on something that needs to be done as soon as I think of it
49	(A40)	Often have the feeling that others laugh or talk about me
50	(E126)	Try to forgive and forget
51	(X244)	Feel little concern for others
52	(H710)	Do things out of revenge
53	(P382)	Have felt contact with a divine power

Appendix A (continued)

No.	IPIP ID	Description
54	(P479)	Am devoted to religion
55	(Q142)	Was a better than average student when I was in school
56	(Q91)	Felt close to my parents when I was a child
57	(A129)	Feel a sense of worthlessness or hopelessness
58	(Q69)	Am interested in science
59	(Q184)	Believe that most questions have one right answer
60	(R36)	Tend to feel happy and irritable at the same time
61	(V62)	Believe it is always better to be safe than sorry
62	(X265)	Rarely get irritated
63	(Q239)	Wanted to run away from home when I was a child
64	(D60)	Often enjoy telling jokes or behaving in a humorous manner
65	(X47)	Make enemies
66	(M38)	Am fascinated by how machines work
67	(X117)	Disliked math in school
68	(X245)	Have a rich vocabulary
69	(H2027)	Like to read
70	(E85)	Don't care about dressing nicely
71	(X5)	Worry about things
72	(H991)	Am easily intimidated
73	(H78)	Love large parties
74	(X135)	Seldom daydream
75	(V316)	Crave the experience of great art
76	(X218)	Tend to vote for liberal political candidates
77	(E115)	Feel sympathy for those who are worse off than myself
78	(V147)	Use my charm to get attention
79	(V172)	Avoid activities that are physically dangerous
80	(D116)	Immediately feel sad when hearing of an unhappy event
81	(D109)	Can make myself work on a difficult task even when I don't feel like trying
82	(Q36)	Pay too little attention to details
83	(V186)	Do not like to visit museums
84	(V335)	Have no special urge to do something original
85	(A132)	Have difficulty showing affection
86	(P407)	Find it easy to manipulate others
87	(V18)	Cannot imagine lying or cheating
88	(X83)	Talk to a lot of different people at parties
89	(P341)	Dislike having authority over others
90	(X82)	Often forget to put things back in their proper place
91	(H784)	Act without consulting others
92	(X104)	Like to visit new places
93	(P450)	Have time for play and relaxation
94	(Q61)	Am open to new experiences
95	(E134)	Don't care what others think
96	(X105)	Am not sure where my life is going
97	(Q165)	Would like to have more power than other people
98	(Q200)	Believe in universal harmony
99	(H1366)	Am skilled in handling social situations
100	(E171)	Try to avoid complex people
101	(P439)	See myself as a good leader
102	(X22)	Avoid crowds
103	(P468)	Am not good at telling jokes
104	(Q11)	Don't use harsh language
105	(M68)	Am able to fix electrical-wiring problems
106	(E90)	Enjoy games of strategy
107	(Q241)	Was a slow learner in school
108	(A142)	Often stop to analyze how I'm feeling
109	(Q51)	Get deeply immersed in music
110	(X201)	Like to solve complex problems

Appendix A (continued)

No.	IPIP ID	Description
111	(H1100)	Am concerned about others
112	(H244)	Like to tidy up
113	(V181)	Don't feel the need to be close to others
114	(Q183)	Love my enemies
115	(V259)	Have an imagination that stretches beyond that of my friends
116	(R65)	Find it difficult to organize tasks and activities
117	(Q215)	Feel that life has no meaning
118	(Q199)	Like being the authority who has everyone's attention
119	(A35)	Have attacked someone physically
120	(V68)	Like to stand out in a crowd
121	(R24)	Get upset if others change the way that I have arranged things
122	(H303)	Want things to proceed according to plan
123	(X84)	Am often in a bad mood
124	(H1203)	Want to be the very best
125	(Q96)	Feel used by other people
126	(X251)	Am able to control my cravings
127	(V295)	Know what to say to make people feel good
128	(E73)	Get so involved with things that I forget the time
129	(A99)	Am often bored
130	(H525)	Demand attention
131	(E11)	Want everything to be "just right"
132	(R71)	Would love to explore strange places
133	(H1093)	Amuse my friends
134	(V308)	Worry about being embarrassed
135	(H660)	Want to be left alone
136	(E57)	Don't know why I do some of the things I do
137	(X163)	Am exacting in my work
138	(X7)	Have difficulty imagining things
139	(X239)	Am not interested in theoretical discussions
140	(P376)	Don't enjoy being the object of jokes
141	(A97)	Plan my life logically
142	(H1119)	Indulge in my fantasies
143	(A56)	Rarely cry during sad movies
144	(X55)	Get to work at once
145	(H976)	Wait for others to lead the way
146	(Q235)	Have never hated anyone
147	(H980)	Need reassurance
148	(R29)	Feel emotions with extreme intensity
149	(H2039)	Spend a lot of time reading
150	(P363)	Don't try to figure myself out
151	(H1018)	Break my promises
152	(H969)	Need a push to get started
153	(V153)	Have a colorful and dramatic way of talking about things
154	(V155)	Come up with new ways to do things
155	(Q240)	Don't like to spend money
156	(V39)	Believe in a universal power or God
157	(H34)	Am the life of the party
158	(D37)	Don't attempt to conform to society's expectations
159	(Q146)	Wonder how I got to be the way that I am
160	(H617)	Feel short-changed in life
161	(X54)	Will not probe deeply into a subject
162	(X64)	Panic easily
163	(Q14)	Have a strong personality
164	(V50)	Prefer to participate fully rather than view life from the sidelines
165	(Q135)	Spend more money than I should
166	(H107)	Make people feel at ease
167	(X87)	Am always prepared
168	(A10)	Keep my feelings to myself, regardless of how unhappy I am
169	(E75)	Spend more money than I have
170	(H1095)	Love to chat
171	(X9)	Am hard to get to know

Appendix A (continued)

No.	IPIP ID	Description
172	(Q154)	Get back at people who insult me
173	(E141)	Am relaxed most of the time
174	(R79)	Check on things more often than necessary
175	(Q23)	Get too tired to do anything
176	(V249)	Don't call attention to myself
177	(V95)	Am a good listener
178	(A3)	Shout or scream when I'm angry
179	(H2013)	Disclose my intimate thoughts
180	(P472)	Am under constant pressure
181	(Q10)	Believe that most people would lie to get ahead

Appendix B

Names, scoring keys and key statistics for the 203 scales contained in the eight inventories approximated by the AMBI. For the scoring keys, numbers correspond to item numbers in Appendix A. Reverse-keyed items are denoted with an R. CC = convergent correlation; α = coefficient Alpha for the GA version and original version of each scale (the latter as reported on the IPIP website).

Scale no.	Inventory	Scale name	Scoring key	CC	α (GA)	α (orig.)
1	NEO PI-R	Anxiety	1R, 71, 162, 147, 33	0.75	0.67	0.83
2	NEO PI-R	Angry hostility	123, 33, 62R, 24, 60	0.69	0.68	0.8
3	NEO PI-R	Depression	57, 33, 160, 72, 1R	0.72	0.65	0.85
4	NEO PI-R	Self-consciousness	72, 134, 49, 57, 30	0.64	0.69	0.74
5	NEO PI-R	Impulsiveness	2, 126R, 136, 165, 14R	0.58	0.56	0.72
6	NEO PI-R	Vulnerability	57, 162, 33, 72, 147	0.66	0.65	0.79
7	NEO PI-R	Warmth	6, 170, 37R, 166, 171R	0.72	0.63	0.8
8	NEO PI-R	Gregariousness	3, 73, 37R, 135R, 102R	0.75	0.7	0.8
9	NEO PI-R	Assertiveness	4, 101, 163, 145R, 21R	0.78	0.74	0.8
10	NEO PI-R	Activity	5, 22, 31R, 101, 163	0.62	0.65	0.72
11	NEO PI-R	Excitement-seeking	44, 46, 106, 79R, 73	0.56	0.39	0.64
12	NEO PI-R	Positive emotions	6, 99, 123R, 133, 37R	0.65	0.59	0.81
13	NEO PI-R	Fantasy	74R, 138R, 142, 115, 59R	0.67	0.55	0.82

(continued on next page)

Appendix B (continued)

Scale no.	Inventory	Scale name	Scoring key	CC	α (GA)	α (orig.)
14	NEO PI-R	Aesthetics	7R, 75, 109, 26R, 27	0.79	0.7	0.84
15	NEO PI-R	Feelings	19, 85R, 179, 150R, 148	0.6	0.59	0.75
16	NEO PI-R	Actions	42R, 27, 94, 92, 138R	0.57	0.6	0.64
17	NEO PI-R	Ideas	110, 139R, 161R, 68, 100R	0.72	0.69	0.82
18	NEO PI-R	Values	76, 54R, 59R, 156R, 98	0.65	0.47	0.78
19	NEO PI-R	Trust	8R, 123R, 160R, 166, 125R	0.57	0.61	0.84
20	NEO PI-R	Straightforwardness	9R, 52R, 86R, 51R, 46R	0.52	0.47	0.74
21	NEO PI-R	Altruism	10, 111, 123R, 65R, 166	0.57	0.58	0.72
22	NEO PI-R	Compliance	104, 172R, 178R, 65R, 34R	0.62	0.6	0.73
23	NEO PI-R	Modesty	11R, 118R, 120R, 38R, 176	0.54	0.76	0.75
24	NEO PI-R	Tender-mindedness	12, 77, 111, 114, 177	0.58	0.49	0.61
25	NEO PI-R	Competence	167, 57R, 116R, 137, 101	0.58	0.53	0.7
26	NEO PI-R	Order	13, 112, 90R, 14, 137	0.69	0.67	0.74
27	NEO PI-R	Dutifulness	151R, 14, 45R, 41, 116R	0.41	0.34	0.67
28	NEO PI-R	Achievement striving	25, 14, 4, 48, 145R	0.62	0.6	0.67
29	NEO PI-R	Self-discipline	14, 116R, 152R, 81, 167	0.64	0.53	0.8
30	NEO PI-R	Deliberation	15R, 35R, 136R, 41, 141	0.55	0.62	0.7
31	HEXACO-PI	Sincerity	9R, 78R, 17R, 86R, 51R	0.52	0.53	0.74
32	HEXACO-PI	Fairness	16R, 46R, 87, 104, 32R	0.54	0.52	0.78

Appendix B (continued)

Scale no.	Inventory	Scale name	Scoring key	CC	α (GA)	α (orig.)
33	HEXACO-PI	Greed avoidance	17R, 97R, 118R, 78R, 16R	0.54	0.66	0.81
34	HEXACO-PI	Modesty	11R, 118R, 97R, 120R, 78R	0.64	0.75	0.8
35	HEXACO-PI	Fearfulness	79, 105R, 164R, 162, 132R	0.67	0.49	0.79
36	HEXACO-PI	Anxiety	1R, 71, 162, 173R, 134	0.71	0.71	0.81
37	HEXACO-PI	Dependence	18, 113R, 147, 3, 168R	0.56	0.53	0.78
38	HEXACO-PI	Sentimentality	19, 80, 143R, 113R, 127	0.74	0.57	0.79
39	HEXACO-PI	Expressiveness	20, 153, 163, 30R, 157	0.76	0.75	0.85
40	HEXACO-PI	Social boldness	21R, 30R, 101, 163, 38	0.85	0.73	0.84
41	HEXACO-PI	Sociability	3, 37R, 88, 170, 73	0.72	0.66	0.81
42	HEXACO-PI	Liveliness	22, 31R, 6, 30R, 99	0.69	0.6	0.79
43	HEXACO-PI	Forgiveness	114, 50, 34R, 23, 121R	0.62	0.58	0.84
44	HEXACO-PI	Gentleness	23, 24R, 104, 40R, 65R	0.66	0.63	0.79
45	HEXACO-PI	Flexibility	24R, 40R, 23, 121R, 172R	0.55	0.64	0.67
46	HEXACO-PI	Patience	24R, 23, 62, 172R, 104	0.69	0.68	0.8
47	HEXACO-PI	Organization	13, 90R, 116R, 112, 14	0.76	0.66	0.87
48	HEXACO-PI	Diligence	25, 81, 48, 14, 116R	0.66	0.6	0.79
49	HEXACO-PI	Perfectionism	82R, 137, 41, 14, 174	0.59	0.51	0.73
50	HEXACO-PI	Prudence	41, 15R, 141, 14, 167	0.57	0.63	0.77

Appendix B (continued)

Scale no.	Inventory	Scale name	Scoring key	CC	α (GA)	α (orig.)
51	HEXACO-PI	Aesthetic appreciation	26R, 83R, 75, 27, 7R	0.75	0.67	0.81
52	HEXACO-PI	Inquisitiveness	27, 58, 83R, 139R, 68	0.72	0.55	0.79
53	HEXACO-PI	Creativity	84R, 115, 138R, 154, 68	0.73	0.63	0.81
54	HEXACO-PI	Unconventionality	28, 139R, 76, 115, 100R	0.74	0.61	0.79
55	JPI-R	Complexity	139R, 76, 43, 100R, 59R	0.62	0.61	0.66
56	JPI-R	Breadth of interest	27, 43, 75, 26R, 83R	0.64	0.69	0.82
57	JPI-R	Innovation	138R, 154, 84R, 115, 110	0.76	0.66	0.88
58	JPI-R	Tolerance	42R, 34R, 121R, 50, 59R	0.55	0.46	0.65
59	JPI-R	Empathy	19, 85R, 113R, 143R, 148	0.59	0.66	0.76
60	JPI-R	Anxiety	1R, 71, 33, 162, 173R	0.73	0.73	0.83
61	JPI-R	Cooperativeness	29, 47R, 95R, 158R, 162	0.66	0.64	0.79
62	JPI-R	Sociability	3, 37R, 135R, 170, 113R	0.67	0.7	0.82
63	JPI-R	Social confidence	30R, 99, 101, 21R, 88	0.76	0.71	0.87
64	JPI-R	Energy level	31R, 22, 175R, 152R, 164	0.64	0.67	0.78
65	JPI-R	Social astuteness	86, 9, 78, 127, 46	0.49	0.52	0.66
66	JPI-R	Risk taking	16, 79R, 46, 61R, 44	0.64	0.45	0.84
67	JPI-R	Organization	13, 14, 116R, 90R, 48	0.66	0.65	0.73
68	JPI-R	Traditional values	76R, 54, 156, 104, 142R	0.68	0.59	0.79
69	JPI-R	Responsibility	32R, 87, 16R, 104, 46R	0.38	0.52	0.66

Appendix B (continued)

Scale no.	Inventory	Scale name	Scoring key	CC	α (GA)	α (orig.)
70	MPQ	Well-being	117R, 6, 22, 57R, 160R	0.63	0.58	0.9
71	MPQ	Social potency	118, 38, 101, 4, 86	0.71	0.75	0.89
72	MPQ	Achievement	25, 110, 81, 137, 180	0.63	0.54	0.84
73	MPQ	Social closeness	3, 37R, 135R, 170, 113R	0.72	0.7	0.86
74	MPQ	Stress reaction	33, 71, 1R, 57, 49	0.73	0.71	0.89
75	MPQ	Aggression	34, 52, 119, 24, 16	0.47	0.47	0.72
76	MPQ	Alienation	49, 125, 57, 160, 180	0.56	0.69	0.82
77	MPQ	Control	35R, 15R, 141, 41, 167	0.61	0.57	0.83
78	MPQ	Harm-avoidance	79, 16R, 132R, 61, 35R	0.5	0.35	0.82
79	MPQ	Traditionalism	76R, 54, 61, 59, 100	0.73	0.52	0.87
80	MPQ	Absorption	36, 109, 75, 108, 28	0.6	0.59	0.9
81	6-FPQ	Affiliation	37R, 88, 3, 170, 6	0.69	0.61	0.78
82	6-FPQ	Dominance	89R, 101, 4, 118, 86	0.75	0.76	0.86
83	6-FPQ	Exhibition	38, 120, 30R, 157, 37R	0.77	0.7	0.8
84	6-FPQ	Abasement	39R, 121R, 140R, 172R, 114	0.4	0.39	0.54
85	6-FPQ	Even-tempered	62, 178R, 24R, 104, 33R	0.62	0.65	0.65
86	6-FPQ	Good-natured	40R, 24R, 172R, 121R, 174R	0.41	0.55	0.58
87	6-FPQ	Cognitive structure	41, 122, 141, 15R, 131	0.54	0.51	0.56
88	6-FPQ	Deliberativeness	15R, 41, 141, 35R, 167	0.63	0.57	0.68
89	6-FPQ	Order	13, 90R, 14, 116R, 82R	0.72	0.63	0.78

(continued on next page)

Appendix B (continued)

Scale no.	Inventory	Scale name	Scoring key	CC	α (GA)	α (orig.)
90	6-FPQ	Autonomy	3R, 135, 28, 95, 113	0.6	0.54	0.59
91	6-FPQ	Individualism	29R, 47, 17R, 95, 120R	0.64	0.47	0.74
92	6-FPQ	Self reliance	91, 70, 170R, 18R, 113	0.46	0.42	0.57
93	6-FPQ	Change	42R, 92, 94, 110, 132	0.52	0.58	0.63
94	6-FPQ	Understanding	43, 27, 69, 68, 139R	0.69	0.65	0.74
95	6-FPQ	Breadth of interest	27, 26R, 83R, 75, 139R	0.68	0.62	0.69
96	6-FPQ	Achievement	25, 81, 48, 152R, 167	0.46	0.58	0.47
97	6-FPQ	Endurance	81, 25, 110, 137, 82R	0.46	0.58	0.59
98	6-FPQ	Seriousness	35R, 93R, 142R, 44R, 94R	0.48	0.45	0.61
99	TCI	Exploratory excitability	44, 94, 92, 42R, 84R	0.67	0.58	0.72
100	TCI	Impulsiveness	35, 41R, 15, 141R, 131R	0.56	0.52	0.75
101	TCI	Extravagance	45, 155R, 165, 169, 15	0.69	0.68	0.83
102	TCI	Disorderliness	46, 9, 16, 86, 104R	0.51	0.52	0.68
103	TCI	Worry & pessimism	1R, 123, 71, 173R, 57	0.69	0.7	0.8
104	TCI	Fear of uncertainty	79, 162, 164R, 1R, 132R	0.68	0.48	0.75
105	TCI	Shyness with strangers	30, 88R, 99R, 37, 21	0.69	0.71	0.87
106	TCI	Fatigability & asthenia	31, 22R, 175, 152, 167R	0.61	0.64	0.85
107	TCI	Sentimentality	19, 143R, 80, 127, 166	0.65	0.54	0.71
108	TCI	Warm communication	37R, 171R, 88, 6, 85R	0.73	0.67	0.86

Appendix B (continued)

Scale no.	Inventory	Scale name	Scoring key	CC	α (GA)	α (orig.)
109	TCI	Attachment	85R, 168R, 171R, 179, 20	0.74	0.69	0.86
110	TCI	Dependence	47R, 95R, 70R, 28R, 158R	0.54	0.61	0.58
111	TCI	Eagerness of effort	48, 144, 152R, 81, 22	0.58	0.64	0.84
112	TCI	Work hardened	137, 66, 81, 110, 144	0.53	0.47	0.75
113	TCI	Ambitious	25, 4, 124, 145R, 48	0.62	0.56	0.79
114	TCI	Perfectionist	25, 14, 81, 180, 137	0.6	0.57	0.76
115	TCI	Responsibility	49R, 57R, 125R, 123R, 1	0.53	0.65	0.78
116	TCI	Purposefulness	96R, 117R, 57R, 101, 164	0.53	0.58	0.77
117	TCI	Resourcefulness	145R, 161R, 72R, 1, 164	0.63	0.57	0.72
118	TCI	Self-acceptance	97R, 17R, 9R, 49R, 131R	0.46	0.43	0.82
119	TCI	Enlightened second nature	126, 136R, 165R, 33R, 14	0.5	0.54	0.84
120	TCI	Social acceptance	50, 40R, 24R, 77, 11R	0.52	0.59	0.77
121	TCI	Empathy	19, 127, 10, 177, 77	0.62	0.65	0.67
122	TCI	Helpfulness	51R, 77, 34R, 123R, 60R	0.39	0.57	0.64
123	TCI	Compassion	52R, 34R, 50, 172R, 114	0.63	0.63	0.88
124	TCI	Pure-hearted conscience	53, 51R, 54, 9R, 21R	0.44	0.42	0.58
125	TCI	Self-forgetful	128, 115, 98, 36, 148	0.66	0.54	0.79
126	TCI	Transpersonal identification	98, 19, 148, 6, 12	0.58	0.6	0.77

Appendix B (continued)

Scale no.	Inventory	Scale name	Scoring key	CC	α (GA)	α (orig.)
127	TCI	Spiritual acceptance	53, 156, 54, 114, 6	0.8	0.73	0.9
128	TCI	Enlightened	53, 54, 156, 114, 6	0.83	0.73	0.95
129	TCI	Idealistic	54, 156, 53, 114, 6	0.76	0.73	0.82
130	CPI	Dominance	21R, 101, 4, 30R, 163	0.78	0.76	0.85
131	CPI	Capacity for status	30R, 21R, 99, 38, 72R	0.51	0.64	0.69
132	CPI	Sociability	99, 30R, 21R, 37R, 101	0.68	0.7	0.78
133	CPI	Social presence	99, 38, 30R, 133, 72R	0.63	0.58	0.73
134	CPI	Self-acceptance	21R, 101, 30R, 163, 99	0.69	0.74	0.65
135	CPI	Independence	72R, 21R, 162R, 57R, 71R	0.63	0.59	0.7
136	CPI	Empathy	99, 30R, 21R, 100R, 133	0.56	0.58	0.64
137	CPI	Responsibility	55, 129R, 60R, 104, 32R	0.36	0.4	0.72
138	CPI	Socialization	56, 63R, 33R, 46R, 104	0.56	0.53	0.72
139	CPI	Self-control	46R, 130R, 33R, 15R, 136R	0.52	0.6	0.79
140	CPI	Good impression	33R, 136R, 169R, 129R, 24R	0.49	0.44	0.78
141	CPI	Communality	31R, 165R, 58, 93, 162R	0.16	0.24	N/A
142	CPI	Well-being	33R, 57R, 160R, 159R, 49R	0.6	0.59	0.8
143	CPI	Tolerance	129R, 60R, 49R, 181R, 121R	0.44	0.52	0.71
144	CPI	Achievement via conformance	55, 15R, 57R, 96R, 60R	0.48	0.43	0.69

Appendix B (continued)

Scale no.	Inventory	Scale name	Scoring key	CC	α (GA)	α (orig.)
145	CPI	Achievement via independence	100R, 49R, 139R, 61R, 58	0.53	0.42	0.75
146	CPI	Intellectual efficiency	57R, 68, 58, 160R, 55	0.64	0.45	0.73
147	CPI	Psychological-mindedness	58, 57R, 49R, 160R, 68	0.48	0.5	0.55
148	CPI	Flexibility	59R, 131R, 13R, 14R, 122R	0.58	0.55	0.72
149	CPI	Femininity	105R, 66R, 79, 162, 89	0.59	0.52	0.66
150	CPI	Vector 1	101R, 118R, 21, 38R, 4R	0.79	0.72	0.85
151	CPI	Vector 2	56, 14, 48, 169R, 104	0.49	0.37	0.74
152	CPI	Vector 3	60R, 49R, 174R, 100R, 160R	0.5	0.46	0.85
153	CPI	Managerial potential	57R, 129R, 21R, 33R, 60R	0.58	0.49	0.79
154	CPI	Work orientation	57R, 160R, 165R, 159R	0.53	0.49	0.73
155	CPI	Creative temperament	61R, 100R, 131R, 138R, 174R	0.5	0.45	0.76
156	CPI	Leadership	57R, 72R, 21R, 96R, 101	0.71	0.6	0.87
157	CPI	Amicability	33R, 123R, 160R, 60R, 56	0.55	0.54	0.79
158	CPI	Law enforcement orientation	14, 137, 116R, 152R, 101	0.35	0.55	0.53
159	CPI	Tough-mindedness	57R, 72R, 136R, 33R, 49R	0.59	0.67	0.74
160	HPI	Empathy	62, 24R, 173, 123R, 23	0.58	0.63	0.62
161	HPI	Not anxious	1, 71R, 33R, 173, 162R	0.67	0.73	0.82

(continued on next page)

Appendix B (continued)

Scale no.	Inventory	Scale name	Scoring key	CC	α (GA)	α (orig.)
162	HPI	No guilt	57R, 49R, 33R, 160R, 159R	0.56	0.59	0.69
163	HPI	Calmness	33R, 57R, 162R, 178R, 148R	0.55	0.56	0.44
164	HPI	Even-tempered	33R, 123R, 62, 24R, 60R	0.54	0.68	0.59
165	HPI	No somatic complaints	31R, 22, 57R, 175R, 71R	0.47	0.67	0.56
166	HPI	Trusting	8R, 123R, 125R, 181R, 60R	0.48	0.58	0.56
167	HPI	Good attachment	63R, 56, 169R, 65R, 151R	0.59	0.48	0.78
168	HPI	Competitive	101, 164, 145R, 25, 89R	0.45	0.69	0.55
169	HPI	Self confidence	72R, 57R, 101, 134R, 145R	0.58	0.68	0.6
170	HPI	No depression	57R, 96R, 33R, 160R, 123R	0.61	0.71	0.78
171	HPI	Leadership	101, 4, 89R, 118, 163	0.76	0.77	0.86
172	HPI	Identity	96R, 57R, 49R, 41, 117R	0.39	0.58	0.82
173	HPI	No social anxiety	21R, 30R, 99, 101, 72R	0.7	0.7	0.75
174	HPI	Likes parties	73, 3, 88, 44, 37R	0.64	0.67	0.67
175	HPI	Likes crowds	102R, 73, 3, 37R, 44	0.5	0.66	0.8
176	HPI	Experience-seeking	94, 132, 44, 27, 42R	0.6	0.58	0.68
177	HPI	Exhibitionistic	38, 176R, 120, 78, 130	0.61	0.8	0.74

Appendix B (continued)

Scale no.	Inventory	Scale name	Scoring key	CC	α (GA)	α (orig.)
178	HPI	Entertaining	64, 103R, 133, 157, 153	0.71	0.74	0.67
179	HPI	Easy to live with	50, 65R, 23, 34R, 6	0.42	0.59	0.55
180	HPI	Sensitive	19, 177, 10, 51R, 127	0.36	0.67	0.29
181	HPI	Caring	19, 166, 177, 127, 170	0.48	0.59	0.43
182	HPI	Likes people	37R, 171R, 3, 88, 170	0.62	0.68	0.75
183	HPI	No hostility	34R, 50, 123R, 172R, 40R	0.49	0.63	0.43
184	HPI	Moralistic	48, 146, 87, 144, 69R	0.4	0.37	0.46
185	HPI	Mastery	14, 122, 48, 25, 137	0.41	0.61	0.34
186	HPI	Virtuous	65R, 104, 23, 8R, 146	0.43	0.56	0.34
187	HPI	Not autonomous	47R, 95R, 29, 147, 158R	0.59	0.68	0.7
188	HPI	Not spontaneous	36R, 148R, 98R, 46R, 142R	0.19	0.38	0.36
189	HPI	Impulse control	35R, 46R, 15R, 16R, 142R	0.6	0.54	0.6
190	HPI	Avoids trouble	158R, 46R, 28R, 181R, 49R	0.37	0.45	0.53
191	HPI	Science ability	58, 105, 66, 110, 138R	0.6	0.62	0.69
192	HPI	Curiosity	66, 105, 58, 110, 95	0.6	0.6	0.64
193	HPI	Thrill-seeking	79R, 132, 46, 44, 106	0.46	0.45	0.65
194	HPI	Intellectual games	106, 110, 176	0.33	0.22	0.5
195	HPI	Generates ideas	101, 138R, 157, 103R, 153	0.55	0.59	0.67

Appendix B (continued)

Scale no.	Inventory	Scale name	Scoring key	CC	α (GA)	α (orig.)
196	HPI	Culture	7R, 83R, 27, 139R, 26R	0.6	0.6	0.59
197	HPI	Education	55, 107R, 68, 69, 61R	0.58	0.67	0.77
198	HPI	Math ability	67R, 106, 110	0.6	0.44	0.78
199	HPI	Good memory	68, 107R, 55, 69, 82R	0.51	0.65	0.54
200	HPI	Reading	69, 149, 43, 68, 107R	0.75	0.77	0.71
201	HPI	Self focus	108, 150R, 159, 49, 33	0.52	0.39	0.69
202	HPI	Impression management	49, 33, 29, 57, 46	0.43	0.56	0.51
203	HPI	Appearance	70R, 95R, 13, 28R	0.43	0.54	0.44

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jrp.2010.01.002.

References

- Ames, D. R., Rose, P., & Anderson, C. P. (2006). The NPI-16 as a short measure of narcissism. *Journal of Research in Personality*, 40(4), 440–450.
- Armstrong, R. D., Jones, D. H., & Wu, I. L. (1992). An automated test development of parallel tests from a seed test. *Psychometrika*, 57(2), 271–288.
- Boyle, G. J. (1991). Does item homogeneity indicate internal consistency or item redundancy in psychometric scales. *Personality and Individual Differences*, 12(3), 291–294.
- Budescu, D. V., & Rodgers, J. L. (1981). Corrections for spurious influences on correlations between MMPI scales. *Multivariate Behavioral Research*, 16(4), 483–497.
- Cattell, R. B., & Tsujioka, B. (1964). The importance of factor-trueness and validity, versus homogeneity and orthogonality, in Test Scales. *Educational and Psychological Measurement*, 24(1), 3.
- Cloninger, C. R. (1994). *The temperament and character inventory (TCI): A guide to its development and use*. Washington University: Center for Psychobiology of Personality.
- Cloninger, C., Przybeck, T., Svrakic, D., & Wetzler, R. (1994). *TCI: The temperament and character inventory: A guide to its development and use*. St Louis (MO): Center for Psychobiology of Personality, Washington University.
- Costa, P. T., Jr., & McCrae, R. R. (1988). Personality in adulthood: A six-year longitudinal study of self-reports and spouse ratings on the NEO personality inventory. *Journal of Personality and Social Psychology*, 54(5), 853.
- Costa, P. T., & McCrae, R. R. (1992). Revised NEO personality inventory (NEO PI-R) and NEO five-factor inventory (NEO-FFI). *Professional Manual: Psychological Assessment Resources*, 297–334.
- Cox, R. M., & Alexander, G. C. (1995). The abbreviated profile of hearing aid benefit. *Ear and Hearing*, 16(2), 176.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334.
- Davis, L. (1991). *Handbook of genetic algorithms*. Citeseer.
- Donnellan, M. B., Oswald, F. L., Baird, B. M., & Lucas, R. E. (2006). The mini-IPIP scales: Tiny-yet-effective measures of the Big Five factors of personality. *Psychological Assessment*, 18(2), 12.
- Francis, L. J., Brown, L. B., & Philipchalak, R. (1992). The development of an abbreviated form of the revised Eysenck personality questionnaire (EPQR-A): Its use among students in England, Canada, the USA and Australia. *Personality and Individual Differences*, 13(4), 443–449.
- Funder, D. C., & Sneed, C. D. (1993). Behavioral manifestations of personality: An ecological approach to judgmental accuracy. *Journal of Personality and Social Psychology*, 64(3), 479–490.
- Gibson, W. M., & Weiner, J. A. (1998). Generating random parallel test forms using CTT in a computer-based environment. *Journal of Educational Measurement*, 35(4), 297–310.
- Goldberg, L. R. (1972). Some recent trends in personality assessment. *Journal of Personality Assessment*, 36(6), 547.
- Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. *Personality Psychology in Europe*, 7, 7–28.
- Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., et al. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality*, 40(1), 84–96.
- Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, 37(6), 504–528.
- Gough, H. G. (2000). *The California psychological inventory. Testing and Assessment in Counseling Practice*, 45, 71.
- Gough, H. G., & Bradley, P. (1996). *California personality inventory manual* (3rd ed.). Palo Alto, CA: Consulting Psychologists Press.
- Gruza, R. A., & Goldberg, L. R. (2007). The comparative validity of 11 modern personality inventories: Predictions of behavioral acts, informant reports, and clinical indicators. *Journal of Personality Assessment*, 89(2), 167–187.
- Hansenne, M., Delhez, M., & Cloninger, C. R. (2005). Psychometric properties of the temperament and character inventory-revised (TCI-R) in a Belgian sample. *Journal of Personality Assessment*, 85(1), 40–49.
- Helson, R., & Moane, G. (1987). Personality change in women from college to midlife. *Journal of Personality and Social Psychology*, 53(1), 176–186.
- Hogan, R., & Hogan, J. (1995). *Hogan personality inventory manual* (2nd ed.). Tulsa, OK: Hogan Assessment Systems.
- Jackson, D. N. (1994). *Jackson personality inventory-revised*. Port Huron, MI: Manual: Sigma Assessment Systems.
- Jackson, D. N., Paunonen, S. V., & Tremblay, P. F. (2000). Six factor personality questionnaire manual. Port Huron, MI: Sigma assessment systems. About feeding methods. *Journal of the American Dietetic Association*, 97(11), 1313–1316.
- John, O. P., Donahue, E. M., & Kentle, R. L. (1991). *The Big five inventory—versions 4a and 5a*. Institute of Personality and Social Research, Berkeley: University of California, Berkeley.
- Kaplan, R. M., & Saccuzzo, D. P. (1982). *Psychological testing: Principles, applications, and issues*. Monterey, CA: Brooks/Cole.
- Kline, P. (2000). *The handbook of psychological testing* (2nd ed.). London, UK: Routledge.
- Lang, A. J., & Stein, M. B. (2005). An abbreviated PTSD checklist for use as a screening instrument in primary care. *Behaviour Research and Therapy*, 43(5), 585–594.
- Langford, P. H. (2003). A one-minute measure of the Big Five? Evaluating and abridging Shafer's (1999a) Big Five markers. *Personality and Individual Differences*, 35(5), 1127–1140.
- Lee, K., & Ashton, M. C. (2004). Psychometric properties of the HEXACO personality inventory. *Multivariate Behavioral Research*, 39(2), 329–358.
- Loevinger, J. (1954). The attenuation paradox in test theory. *Psychological Bulletin*, 51(5), 493–504.
- McClelland, D. C. (1980). Motive dispositions: The merits of operant and respondent measures. *Review of Personality and Social Psychology*, 1, 10–41.
- McCrae, R. R. (2001). Traits through time. *Psychological Inquiry*, 85, 87.
- McCrae, R. R., & Costa, P. T. (2007). Brief versions of the NEO-PI-3. *Journal of Individual Differences*, 28(3), 116.
- McDonald, R. P. (1981). The dimensionality of tests and items. *British Journal of Mathematical and Statistical Psychology*, 34(1), 100–117.
- Mitchell, M. (1998). *An introduction to genetic algorithms*. The MIT Press.
- Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.
- Osburn, H. G. (2000). Coefficient alpha and related internal consistency reliability coefficients. *Psychological Methods*, 5(3), 343–355.
- Oswald, F. L., Friede, A. J., Schmitt, N., Kim, B. H., & Ramsay, L. J. (2005). Extending a practical method for developing alternate test forms using independent sets of items. *Organizational Research Methods*, 8(2), 149.
- Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five inventory in English and German. *Journal of Research in Personality*, 41(1), 203–212.
- Roberts, B. W., & DelVecchio, W. F. (2000). The rank-order consistency of personality traits from childhood to old age: A quantitative review of longitudinal studies. *Psychological Bulletin*, 126(1), 3–25.
- Robins, R. W., Hendin, H. M., & Trzesniewski, K. H. (2001). Measuring global self-esteem: Construct validation of a single-item measure and the Rosenberg self-esteem scale. *Personality and Social Psychology Bulletin*, 27(2), 151.
- Saucier, G. (1994). Mini-markers: A brief version of Goldberg's unipolar Big-Five markers. *Journal of Personality Assessment*, 63(3), 506–516.
- Schuerger, J. M., Zarella, K. L., & Hotz, A. S. (1989). Factors that influence the temporal stability of personality by questionnaire. *Journal of Personality and Social Psychology*, 56(5), 777–783.
- Smith, G. T., McCarthy, D. M., & Anderson, K. G. (2000). On the sins of short-form development. *Psychological Assessment*, 12(1), 102–111.
- Stanton, J. M. (2000). Empirical distributions of correlations as a tool for scale reduction. *Behavior Research Methods, Instruments, & Computers: A Journal of the Psychonomic Society Inc.*, 32(3), 403.
- Stanton, J. M., Sinar, E. F., Balzer, W. K., & Smith, P. C. (2002). Issues and strategies for reducing the length of self-report scales. *Personnel Psychology*, 55(1), 167–194.

- Team, R. D. C. (2008). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Tellegen, A., & Waller, N. G. (1994). Exploring personality through test construction: Development of the multidimensional personality questionnaire. *Personality Measures: Development and Evaluation*, 1, 133–161.
- Troidahl, V. C., & Powell, F. A. (1965). A short-form Dogmatism scale for use in field studies. *Social Forces*, 44, 211.
- Usala, P. D., & Hertzog, C. (1991). Evidence of differential stability of state and trait anxiety in adults. *Journal of Personality and Social Psychology*, 60(3), 471–479.
- Whitley, D. (1994). A genetic algorithm tutorial. *Statistics and Computing*, 4(2), 65–85.
- Willighagen, E. (2005). Genalg: R based genetic algorithm. *R package version 0.1*, 1.