Pergamon

Learning and Individual Differences
12 (2000) 81–103

Learning and
Individual Differences

The Armed Services Vocational Aptitude Battery (ASVAB) Little more than acculturated learning (Gc)!?☆

Richard D. Roberts[a,]*, Ginger Nelson Goff[b], Fadi Anjoul[a], P.C. Kyllonen[c], Gerry Pallier[a], Lazar Stankov[a]

[a]*Department of Psychology, The University of Sydney, Sydney, NSW, Australia*
[b]*Metrica Inc., San Antonio, TX, USA*
[c]*Center for New Constructs, Educational Testing Services, Princeton, NJ, USA*

Abstract

The Armed Services Vocational Aptitude Battery (ASVAB) is administered to over 1 million participants in the USA each year, serving either as a screening test for military enlistees or as a guidance counseling device in high schools. In this paper, we examine the factorial composition of the ASVAB in relation to the theory of fluid and crystallized intelligence and Carroll's [1993. *Human cognitive abilities: a survey of factor-analytic studies*. New York: Cambridge Univ. Press.] three-stratum model. In two studies ($N=349$, $N=6751$), participants were administered both the ASVAB and tests designed to measure factors underlying these (largely) analogous models. Exploratory and confirmatory factor analyses (CFA) of correlational data suggested that the ASVAB primarily measures acculturated learning [crystallized intelligence (Gc)]. This evidence does not support the frequent claim that this test measures psychometric *g*. Our conclusion is that the ASVAB should be revised to incorporate the assessment of additional broad cognitive ability factors, particularly fluid intelligence and learning and memory constructs, if it is to maintain its postulated function. © 2001 Elsevier Science Inc. All rights reserved.

Keywords: ASVAB; Psychological testing; Fluid and crystallized intelligence; Personal selection

☆ This research was conducted while the principal investigator held a National Research Council Fellowship at the Human Effectiveness Directorate of the US Air Force Research Laboratory, Brooks AFB, TX, USA. Due acknowledgment is given to all supporting institutions. However, the views expressed herein are those of the authors, and as such, are not intended to reflect official government policy. Part of this paper was presented on October 15, 1997 at the International Military Testing Association Meeting, Swiss Grand Hotel, Sydney, Australia. A further portion was presented on July 5, 1999 at the Ninth Biennial Meeting of the International Society for the Study of Individual Differences, Coast Plaza Hotel, Vancouver, BC, Canada.

* Corresponding author. Fax: +61-2-9351-2603.
E-mail address: richardr@psych.usyd.edu.au (R.D. Roberts).

1041-6080/00/$ – see front matter © 2001 Elsevier Science Inc. All rights reserved.
PII: S1041-6080(00)00035-2

1. Introduction

The resemblance of many of today's 'intelligence' tests to those developed by Binet and Simon (1916), Wechsler (1981), and others (e.g., Yerkes, 1921) testify to the pioneering psychologists' perspicacity and imagination. Indeed, Cronbach (1984, p. 201) draws a favorable analogy to one of the 20th century's most celebrated inventions: 'Tests are like automobiles . . . the main working parts of today's machines were to be found in the cars of 1920 — society is slow to supplant an invention that works.'

The Armed Services Vocational Aptitude Battery (ASVAB; U.S. Department of Defense, 1984) would appear, at least on face value, to be one such psychometric instrument. Consider the following facts, signifying both its importance and permanence. Within the USA, performance on this instrument is a major determinant in the career choices of over 1.3 *million* young women and men *per annum* (Kaplan & Saccuzzo, 1997). In empirical instantiations, numerous psychological studies have been devoted to (or employ) the ASVAB somewhere within their experimental design, generating as it were a multimillion-dollar research industry. For example, the recent controversies engendered by the publication of *The Bell Curve* (Herrnstein & Murray, 1994), with all its accompanying literature, have direct links to the ASVAB. It was, after all, from the 1980 standardization sample of the ASVAB (U.S. Department of Defense, 1982) that much of *The Bell Curve*'s data were computed. The ASVAB also shares an important place in the history of the mental testing movement (see, e.g., Carroll, 1997; Gregory, 1996; Hunt, 1995). Indeed, within contemporary psychology, it has boldly been proclaimed that 'The ASVAB is representative of *the state of the art* in multiple aptitude batteries' (Ree & Carretta, 1995, p. 269, italics added; see also Jensen, 1985; Murphy, 1984).

Despite a pervasive feeling that tests like the ASVAB may serve an important predictive function and by extension contribute to the successful operation of both military and civilian organizations, it is *not* clear to what extent this instrument should remain immutable. Because the understanding of human cognitive abilities presumably remains incomplete, tests should grow cumulatively with the corpus of research knowledge. Anything short of this is antithetical to scientific progress and may ultimately inhibit practical advantages that a psychological test brings to the wider community. A belief in the 'finality' of our models of intelligence and by extension the tests that accompany them is, as Hearnshaw (1951, p. 316) stated 50 years ago:

> [H]euristically inhibitory and contrary to the exploratory genius of scientific research. The scientist is unlikely to discover anything unless s/he believes that there is something to discover. The history of science, moreover, warns us that even in the more mature physical sciences proclamations of finality, even when apparently most firmly founded, have been falsified by the progress of research. It would therefore be rash to claim anything like finality in the psychologist's map of the intellect unless the arguments are logically overwhelming.

1.1. Test construction: a dynamic process

Principles for constructing *psychometrically* sound psychological tests are well established (e.g., Anastasi & Urbina, 1997; Cronbach, 1984; Kaplan & Saccuzzo, 1997; Murphy

& Davidshofer, 1998). Advances in a variety of statistical techniques over the past decade, and especially improvements in item response theory (IRT) and confirmatory factor analysis (CFA), hold still greater promise towards this end. However, there would appear a number of equally consequential theoretical issues within the field of individual differences (and related disciplines) that suggest constant attention be afforded to various subtests (and even items) comprising any test battery. Unless a given test is subjected to this 'review process,' it is unlikely that the instrument will retain either its overall integrity or indeed be of continuing practical utility. Although formal, systematic treatment of the impact of psychological theory on test construction is seldom addressed in the scientific literature (see, however, Matarazzo, 1992 for a notable exception), several factors seem more critical than has been explicitly acknowledged.

One major influence affecting psychometric test construction is the notion that the capabilities indicating human intelligence are themselves changing over time as a function, in particular, of technological and cultural evolution (Horn & Noll, 1994). New capabilities appear with every innovation (e.g., computer proficiency), while competencies that were once very important (e.g., spelling ability) are less so now. This state of affairs may occur either because society no longer requires the underlying capability or technology has rendered it obsolescent. For instance, knowing how to use a slide rule (an attribute once valued) brings few rewards to the modern mathematician, and the word processor's spell-checking tool has made lexical ability less important that it once was. In light of the dynamic nature of acculturated abilities, tests need constantly to be redeveloped and refined to reflect the attributes most valued by the dominant culture.

Arguably, a more serious problem occurs if tests remain static in the face of developments in theories concerning the structure of human intelligence. In an astute appreciation of the consequences of such conservatism, Kaufman (1979, p. 4) lamented that mental testing (in general) had actually failed to

> [G]row conceptually with the advent of important advances in psychology and neurology . . . The impressive findings in the areas of cognitive development, learning theory, and neuropsychology during the past 25–50 years have not invaded the domain of the individual intelligence test. Stimulus materials have been improved and modernized; new test items and pictures have been constructed with keen awareness of the needs and feelings of both minority-group members and women . . . However, both the item content and the structure of intelligence tests have remained basically unchanged.

It might be countered that the importance of making a test *congruent with theory* is merely a cosmetic exercise. However, consider the following. The original test upon which all others are based (The Stanford–Binet Intelligence Scale) has recently gone through its fourth revision (Thorndike, Hagen, & Sattler, 1985). Rather than modernize test items and provide a general IQ score, the authors redeveloped the test to conform to the theory of fluid (Gf) and crystallized (Gc) intelligence. This revision was undoubtedly prompted by the sheer weight of developmental evidence concerning cognitive differentiation and by a need to expand the universe of assessment beyond that of an acculturated (Gc) kind (see Anastasi, 1988). On the other hand, the Wechsler scales (e.g., Wechsler, 1981), viewed by many commentators as the prototypical intelligence test *par excellence*, have remained (aside from item modifications) relatively untouched since their inception (see Frank, 1983). Thus, while the adult version has recently gone through its third revision, it retains the contentious Verbal vs. Performance IQ distinction. Studies of the Wechsler Adult Intelligence Scale (WAIS) have consistently demonstrated that these scales are factorially impure (see, e.g., Carroll, 1993, pp. 701–702; McArdle & Horn, 1983). Indeed, different scoring procedures (rather than the scale scores presented in the manual) are often implemented by clinicians when employing this instrument for assessment purposes (Senior, 1996). It is unlikely such *post-hoc* treatment is as informative to practitioners as would be a complete redevelopment of the test protocol according to some substantive model. The third revision of the WAIS makes some concessions to this possibility but, in our opinion, has not gone far enough (see Pallier, Roberts, & Stankov, 2000). Indeed, assessing processing and trait constructs for different tests in a (largely) arbitrary manner (see McGrew & Flanagan, 1998) blurs important conceptual boundaries.

Thankfully, a trend towards developing tests on the basis of *established* psychological theories is now becoming more commonplace than in the time of Kaufman's (1979) critique (Daniel, 1997). Thus, several new tests have been constructed using contemporary theories of intelligence (see, e.g., Woodcock & Johnson, 1989), often with recourse also to developmental (e.g., Piagetian) and neuropsychological (e.g., Lurian) frameworks (e.g., Kaufman & Kaufman, 1993; Naglieri & Das, 1997). Despite this trend, it would seem an oversight that an APA task force, examining the 'knowns and unknowns' of intelligence (Neisser, 1997; Neisser et al., 1996) failed to give this direction in mental testing any coverage at all (Naglieri, 1997). This issue is actually highly pertinent since, for all the research conducted on the topic, arguably the most frequently cited definition of intelligence remains the euphemistically operational: intelligence is what the tests test (Boring, 1923).

1.2. The theory of fluid and crystallized intelligence

A prominent theory guiding the construction of many recent test batteries is that of fluid and crystallized intelligence (see, e.g., Horn & Cattell, 1967). This model considers that there is enough structure among primary mental abilities to define several distinct types of broad cognitive ability. The model derives its name from the two broad cognitive abilities most extensively studied. The main distinguishing feature between fluid and crystallized intelligence is the amount of formal education and acculturation present in the content of, or operations required during, tests measuring these abilities. It is well established that fluid intelligence (Gf) depends to a much smaller extent on formal education experiences than does crystallized intelligence (Gc; e.g., Horn & Hofer, 1992; Horn & Noll, 1994; Stankov, Boyle, & Cattell, 1995).

The theory of fluid and crystallized intelligence incorporates a number of factors in addition to the ones from which it derives its name. Some, such as broad auditory function (Ga) and broad visualization (Gv), are related to perceptual processes. Further factors, including short-term acquisition and retrieval (SAR) and tertiary storage and retrieval (TSR), are related to memory processes, while others, such as clerical-perceptual speed (Gs), reflect speed in performing tasks of relatively trivial difficulty. Each of these factors is assumed to share differential relations with external measures (such as age), and each is postulated to arise from the workings of different cognitive and neurophysiological functions.

1.3. The three-stratum model

Carroll's (1993) three-stratum model of intelligence shares a number of conceptual parallels with Gf/Gc theory, and in particular, with respect to the level of prominence that is given to second-order constructs. Carroll (1993) arrived at this model after an extensive reanalysis of some 477 data sets collected within the psychometric discipline this century. (These included many of the studies that formed the basis of Gf/Gc theory). Because this model serves to provide a comprehensive taxonomy for current theory, research, and practice involving human cognitive abilities, each of the constructs supported in this reanalysis (and subsequently encapsulated under this model) is represented in Fig. 1. Notably, Carroll (1993) found only one factor having no analogue in Gf/Gc theory — Broad Processing Speed — a construct he also suggests is poorly understood (see, however, Roberts & Stankov, 1999). This factor, notwithstanding the degree of convergence between the three-stratum model and Gf/Gc theory (which preceded the former by at least three decades) is compelling.[1]

1.4. The ASVAB: a critique

The ASVAB is difficult to place *precisely* within any comprehensive theoretical framework.[2] Thus, Carroll's (1993, p. 699) reanalysis indicated that one or more of the subtests comprising the ASVAB measured the following primary (i.e., first stratum) factors: Verbal Ability, Quantitative Reasoning, Numerical Facility, Mechanical Knowledge, Knowledge of Mathematics, Perceptual Speed, and General Information.[3] The main reason the ASVAB was constructed without any obviously coherent factorial structure is quite clear. The initial

---

[1] Carroll (1993) does adopt slightly different terminology for his broad factors — a fact that may be readily observed in Fig. 1. For example, clerical-perceptual speed is designated Broad Speediness, while SAR is conceptualized as Broad Memory and Learning (Gy). For the most part, this represents different nomenclature for very similar constructs. A more noteworthy disparity is in the importance attached to the general factor. In this instance, psychologists subscribing to Gf/Gc theory often cite lack of factorial invariance across test batteries as limiting the generalizability (and interpretability) of a third-order general intelligence construct (see, e.g., Horn, 1985, 1998; Roberts, Pallier, & Goff, 1999). In short, Carroll's model and the theory of fluid and crystallized intelligence are roughly equivalent, especially with respect to the interpretation of first- and second-strata factors.

[2] Ree and Carretta (1994) claim to demonstrate that the factorial structure of the ASVAB is rather similar to a relatively antiquated model of intelligence first put forward by Vernon (1960). It should be noted that even Vernon (1960) assumed that, over time, more than two factors, (spatial/mechanical) and (verbal/educational), would occupy a stratum just below psychometric *g* (see Carroll, 1993, p. 638). This caveat is nowhere acknowledged by Ree and Carretta (1994) nor do they consider more plausible hierarchical models (e.g., Gf/Gc theory). Note also that like the ASVAB, Vernon's (1960) model emanates from research aimed at satisfying military personnel selection requirements. This provides a somewhat narrow basis for a model of intellect. Our argument is that a more comprehensive view allows for a principled selection of tests that may better fit the changing conditions of work and life in a modern society.

[3] Three marker tests are considered minimally acceptable to differentiate between constructs using factor analytic techniques (Carroll, 1993). Thus, without additional reference tests (which fortunately Carroll (1993) had at his disposal), it would not have been clear that any of these factors were necessarily being assessed by the ASVAB.

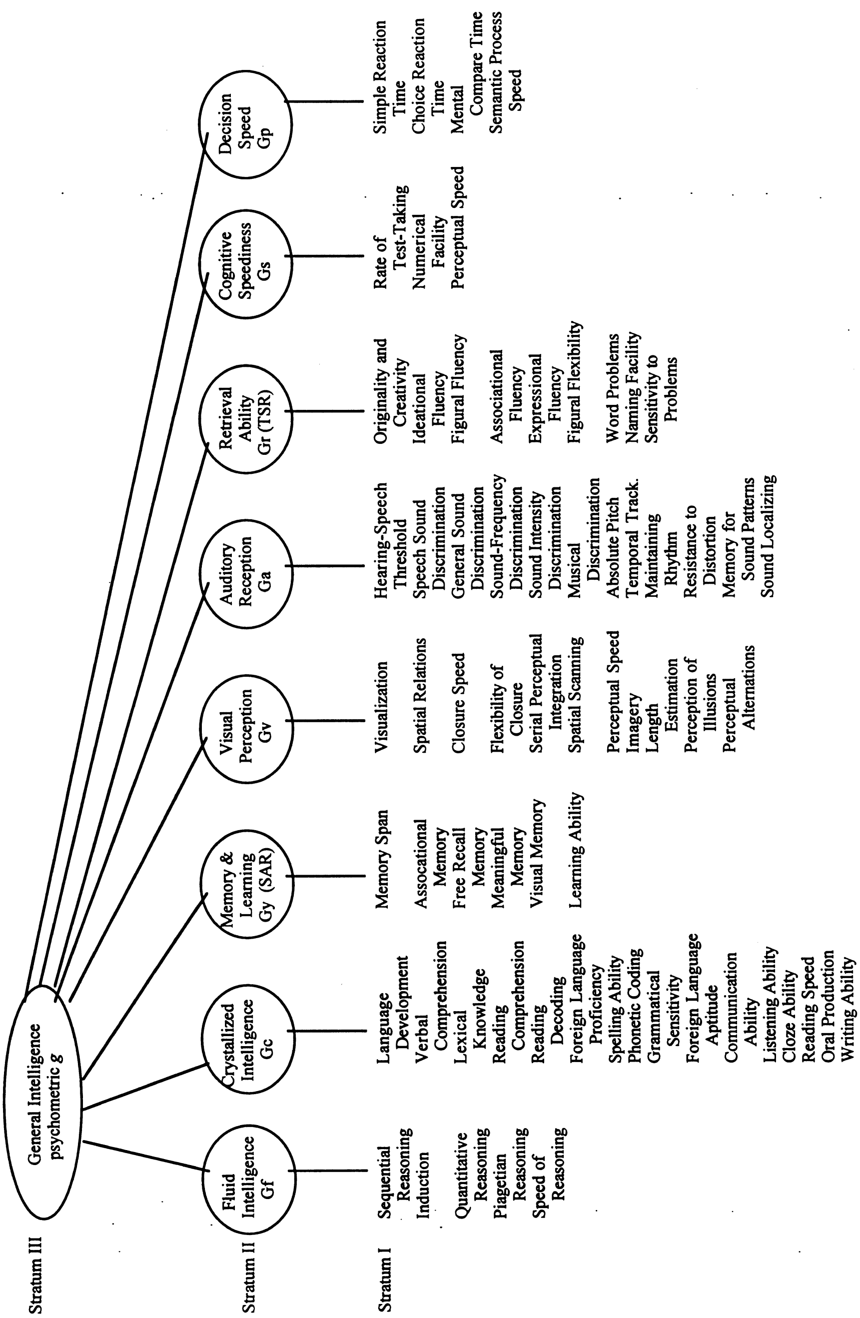


Fig. 1. Carroll's (1993, p. 626) three-stratum model of the structure of human cognitive abilities.

purpose of this multiple-aptitude battery was as a classification instrument. Therefore, tests were selected on the basis of perceived similarities to military occupations rather than any psychological theory. Note also that at the time of its development, the efficacy of several competing models of human cognitive abilities was still contentious.

Nevertheless, it should be emphasized that from the perspective of Gf/Gc theory, the ASVAB would appear (intuitively at least) to be comprised mainly of ability tests that would define crystallized intelligence at a second stratum. [Primary factors that appear as exceptions to this interpretation include Quantitative Reasoning (Gf), Mechanical Knowledge (possibly Gv), and Perceptual Speed (Gs).] In short, it is unclear whether any broad ability factors other than Gc (which is seemingly overdetermined) may be sufficiently defined, since markers of other second-stratum factors are lacking in the battery's design.

The purpose of the present investigation was to examine the ASVAB within the context of Gf/Gc theory, whose broad cognitive ability constructs are generally analogous to Carroll's (1993) second-stratum factors. To present knowledge, although numerous studies have been conducted with the ASVAB, there has been no attempt to ascertain how it relates to factors found at this stratum and in light of these models. For this purpose, both exploratory and CFA were conducted on two independent samples, given both the ASVAB and tests chosen (*a priori*) on the basis of established substantive theory.

The foregoing analyses of the factorial composition of the ASVAB are not merely of practical utility but of conceptual relevance. Recently, Ree and his colleagues have conducted many studies with the ASVAB as the psychometric referent (see, e.g., Ree & Carretta, 1994, 1996; Ree & Earles, 1992, 1993; Ree, Earles, & Teachout, 1994; Stauffer, Ree, & Carretta, 1996). In each instance, it is purportedly demonstrated that the prediction of job performance, training criteria, and the like is 'not much more than *g*.' In this context, *g* appears to be envisaged as 'the first principal component extracted from an infinite battery of psychometric tests' (Jensen, 1979, p. 18). However, in reality, the *g* extracted from the ASVAB may be heavily biased towards Gc (cf. Gustafsson & Muthén, 1994). An empirical demonstration of this possibility would suggest that some important (and relatively self-evident) caveats be drawn to findings made within the research program conducted by Ree and his associates.

It should be emphasized that finding the ASVAB to reflect Gc, and little else, would also pose serious problems for the social policy recommendations made in Herrnstein and Murray's (1994) controversial book, *The Bell Curve*. For, if it is acknowledged that it is only acculturated intelligence that is in decline among the so-called 'underclass' and not psychometric *g per se*, important qualifications would need to be made to many of the observations made in *The Bell Curve*. The current investigation thus has major implications for a range of pertinent issues in basic and applied psychological research.

2. Study 1

2.1. Rationale

The aim of Study 1 was to examine the factor structure of the ASVAB when additional tests, selected on the basis of Gf/Gc theory, were included in the experimental design. A decision to employ several markers in order to overdetermine Gf was based on the conviction that Gf is closest to the general factor (e.g., Carroll, 1993; Marshalek, Lohman, & Snow, 1983). Indeed, some psychometricians have proposed models of intelligence wherein *g* and Gf are one and the same (see Gustafsson, 1984, 1988; Gustafsson & Muthén, 1994). With this in mind, if the ASVAB measures a robust general factor akin to psychometric *g*, tests from this battery should tend either to load highly on the fluid intelligence factor or else have high correlation with this construct at the second order. On the other hand, should tests from the ASVAB (on the main) define a separate factor that shares low correlation with fluid intelligence, then likely it is not the measure of psychometric *g* that advocates of this instrument have proclaimed.

This study is also of relevance to the issue of whether or not it would be expedient to include additional tests in future revisions of the ASVAB. Concern regarding this issue has already been expressed in the literature, with research into the inclusion of an object assembly task already underway (e.g., Sterling, Goff, & Sawin, 1997). Consistent with the arguments presented in the Introduction, it is unclear what effects such 'cosmetic changes' would have on the factorial composition of the test battery. The present analyses were intended to highlight an alternative approach that would suggest tests be selected (or perhaps discarded) *a priori* on the basis of an established substantive model.

2.2. Participants

Participants were 349 (43 female) US Air Force recruits undergoing their sixth week of basic training. The age of the sample ranged from 17 to 31 years, with a mean of 20.20 years (S.D. = 2.20). All individuals scored at or above the 40th percentile on the Armed Forces Qualifying Test. Because of the selection practices employed within the organization, all participants had (minimally) finished high school (or the equivalent) and had their vision corrected to meet Air Force standards.

2.3. Design and procedure

A total of 11 tests demarcating factors within Gf/Gc theory were administered to all participants by a trained test proctor in moderate-sized groups consisting of between 20 and 40 participants. These marker tests are listed in the top section of the first column of Table 2. Following Carroll's (1993) model, Tests 1–4 assess fluid intelligence (Gf), Tests 5–6 are markers for broad visualization (Gv), Tests 7–8 measure SAR, and Tests 9–11 assess the clerical-perceptual speed (Gs) factor.[4] Importantly, each marker test was selected on the basis of past research, demonstrating the clarity a respective test brings to issues of cognitive ability structure (see, e.g., Davies, Stankov, & Roberts, 1998; Roberts, 1997; Roberts, Stankov, Pallier, & Dolph, 1997; Stankov, 1988; Stankov, Seizova-Cajic, & Roberts, 2001).

[4] It should be acknowledged from the outset that it has been known for some time now that there may be difficulties in distinguishing between Gf and Gv abilities, especially where the relations among the visual patterns are not clearly manifested (see, e.g., Humphreys, 1962; Roberts et al., 1997).

Table 1
Latent roots and percentage of variance accounted for by each of the five extracted factors (Study 1)

Latent root	1	2	3	4	5
Eigenvalue	5.4	2.8	1.7	1.4	1.1
% Variance	25.8	13.2	8.0	6.7	5.2

Following informed consent from the Air Force enlistees participating in the investigation, data from the ASVAB were also collected from their military records. This test battery was administered 2–6 months earlier as part of each enlistee's formal Air Force admission requirements. The ASVAB consists of the 10 subtests shown in the first column of the bottom section of Table 2 (i.e., Tests 12–21). Because of their centrality, both in this study and the one that follows, they are also described briefly in Appendix A.

2.4. Results

It was determined using a variety of criteria (i.e., root-one, Scree plot, Montanelli–Humphreys parallel analysis) that five factors should be extracted from the data matrix.[5] The percentage of total test variance accounted for by these five factors (along with latent roots and Eigenvalues) is given in Table 1.

The amount of common variance captured by this first factor (25.8%) is notably smaller than that reported in a number of studies. It seems plausible that this is because of our attempt to sample more representatively across the domain of psychometric indices. The point that a more diverse collection of (nonetheless established) measures will tend to give a weak general factor is often ignored in models postulating the importance of psychometric *g* (see Horn, 1998, who makes a similar point). Using maximum likelihood procedures, these five factors were rotated to an oblique (i.e., oblimin) solution. The results of this analysis [along with the factor intercorrelation matrix and communalities (h^2)] are presented in Table 2.

Inspection of Table 2 reveals a clearly defined fluid intelligence factor. However, only two ASVAB tests (i.e., Math Knowledge and Arithmetic Reasoning) have (moderate) loadings on Factor 1.[6] No tests from the ASVAB load on the SAR factor, a memory factor of some importance, at least in clinical applications, and certainly also relevant to educational attainment. Factor 3, which may be interpreted unequivocally as Gs, shares loadings from tests requiring speed in relatively simple tasks that with unlimited time, participants should, in principle, complete with 100% accuracy. As would be expected, the two ASVAB tests thought to assess this construct do indeed share substantial loadings on the Gs factor. This outcome makes the interpretation of other constructs presented in this analysis more credible than if this had not been the case.

[5] All correlational (or, in the case of CFA, covariance) matrices generating the results reported in this paper are available from the authors upon request.

[6] It should be noted that Horn (1998) has, in recent times, argued that there exists a separate broad quantitative factor denoted as Gq. Apparently, with insufficient markers of this domain, tests assessing this ability will tend to load with Gf. The factorial complexity evidenced in ASVAB tests assessing quantitative aptitude probably reflects this phenomenon.

Table 2
Oblimin factor pattern matrix of psychometric tests (Study 1)

Psychometric tests	F1: Gf	F2: SAR	F3: Gs	F4: Gc	F5: TK	h^2
Selected tests						
1. Progressive Matrices	0.67	0.04	− 0.13	0.08	− 0.02	0.43
2. Letter Sets	0.51	0.09	0.09	0.01	0.06	0.35
3. Number Series	0.40	0.09	0.27	0.01	0.10	0.37
4. Letter Counting	0.26	0.31	− 0.01	− 0.11	− 0.02	0.20
5. Card Rotation	0.35	0.01	0.28	0.01	0.25	0.39
6. Hidden Figures	0.40	0.10	0.17	− 0.02	0.21	0.36
7. Digit Span (Forward)	− 0.11	0.76	0.02	0.11	− 0.04	0.58
8. Digit Span (Backward)	0.11	0.71	− 0.01	− 0.01	0.04	0.56
9. Finding A's	0.16	− 0.15	0.46	− 0.05	− 0.11	0.27
10. Search	0.21	− 0.05	0.60	0.01	0.04	0.48
11. Number Comparison	0.01	0.06	0.67	0.00	0.02	0.47
ASVAB						
12. Coding Speed	− 0.12	0.09	0.73	0.04	0.04	0.54
13. Numerical Operations	− 0.16	0.05	0.69	0.05	− 0.07	0.47
14. General Science	0.17	− 0.06	− 0.09	0.69	0.17	0.66
15. Word Knowledge	0.01	− 0.02	− 0.04	0.78	0.01	0.61
16. Paragraph Comprehension	− 0.07	0.08	0.09	0.45	0.02	0.23
17. Math Knowledge	0.44	0.01	0.14	0.36	− 0.06	0.44
18. Arithmetic Reasoning	0.36	0.10	0.20	0.30	0.13	0.50
19. Mechanical Comprehension	0.31	0.03	− 0.10	0.16	0.59	0.69
20. Electronic Information	0.05	− 0.05	− 0.09	0.24	0.64	0.60
21. Autoshop Information	− 0.14	− 0.01	− 0.01	− 0.06	0.88	0.68
The factor intercorrelation matrix						
Factor	Gf	SAR	Gs	Gc		
SAR	.26	–				
Gs	.31	.18	–			
Gc	.22	.13	.17	–		
TK	.31	.15	.00	.43		

All loadings above 0.30 are underlined.

The two remaining factors given in Table 2 are defined by ASVAB tests. Because of the very clear split between Gf and Factor 4 in this data set, and the high loadings exhibited by tests of a verbal-educational nature, the first of these factors is interpreted as the acculturated learning factor (i.e., Gc).[7] Finally, Factor 5 is nominated a Technical Knowledge (TK) factor, which has small (nonsalient) loadings from the marker tests of broad visualization (Tests 5 and 6). The magnitude of loading is somewhat lower than expected given that visualization is an important component of almost all the TK tests.

[7] In terms of conclusions reached later in the paper, it is worth noting that Math Knowledge and Arithmetic Reasoning share loadings both on this factor and Gf (i.e., these tests are factorially complex). In the absence of any test clearly demarcating fluid intelligence in the ASVAB, the general factor extracted from that battery should undoubtedly be interpreted as broad crystallized intelligence (Gc).

Factor intercorrelations are equally informative. The correlation between Gf and Gc is in the lower range of that reported in the literature. However, it is of similar magnitude to other studies where Gf and Gc are suitably defined (i.e., sufficient markers of each higher-order cognitive ability are employed; see, e.g., Davies et al., 1998; Roberts, 1997; Stankov et al., 2001). This result argues very strongly both for the independence of these two constructs and the fact that the ASVAB under-represents an important cognitive ability. Note also that the magnitudes of all other factor intercorrelations presented in this lower table are consistent with those typically found by researchers working within the framework provided by Gf/Gc theory (Roberts & Stankov, 1999). In short, the tests defining the Gf/Gc constructs in the design behaved in a remarkably lawful manner. However, the majority of ASVAB tests loaded on factors that were distinct from these constructs at both the first and second orders.

2.5. Discussion

The present data question the extent to which the ASVAB provides an adequate assessment of psychometric *g per se*. In fact, this limitation in the ASVAB is highlighted if one considers the fact that an overwhelming body of evidence indicates Gf to be closer to the first general factor than Gc (see, e.g., Carroll, 1993). In a similar vein, the data indicate that two Gf markers (minimally) need to be employed in the factorial composition of the ASVAB if it is to represent this construct adequately. However, it could be objected (conservatively, perhaps even pedantically) that (a) the sample size ($N=349$) is not quite sufficient; (b) the study employs exploratory rather than CFA techniques; and (c) the interpretation of Factor 4 would be more compelling if other Gc measures were included in the design. Moreover, it should be recalled that the ASVAB test scores were collected several months prior to the psychometric tests introduced into the experimental design. Discrepant results might therefore be attributed to artifacts associated with time of testing. A second study is reported that addresses each of these various abovementioned concerns.

3. Study 2

3.1. Rationale

Whilst considering the above issues, the main aim of Study 2 was to investigate the factor structure of the ASVAB when a particularly diverse selection of ancillary tests was included in the experimental design. Thus, this section may be viewed as an attempt to replicate and extend the results presented in Study 1. To achieve this purpose, marker tests from the ETS Kit of Factor-Referenced Cognitive Tests (hereafter referred to as the Kit; Ekstrom, French, Harmon, & Derman, 1976) were given to the same target population (i.e., Air Force enlistees). The Kit tests were designed to capture individual differences in almost all of the first stratum (i.e., primary factors) of human cognitive abilities (Ekstrom et al., 1976). The data used in Study 2 were originally gathered in 1986–1987 by Wothke, Bock, Curran, Fairbank, Augustin, Gillet, & Guerrero (1991) and were reanalyzed for the purposes of the present investigation using CFA techniques.

In principle, the Kit tests provide a structure that is similar to the full-blown model of fluid and crystallized intelligence discussed in the Introduction to this paper (see also, Carroll, 1993). Indeed, the only construct not assessed is broad auditory function (Ga) — largely because the Kit relies exclusively on the traditional paper-and-pencil test format. Thus, the Kit includes multiple markers for the following second-order factors: Gf, Gc, SAR, broad visualization (Gv), TSR, and clerical-perceptual speed (Gs). Should the ASVAB provide a fallible index of psychometric *g*, then most tests comprising that battery should load on the Gf factor and/or have substantial loadings on a factor that is highly correlated with Gf.

This second study is also relevant to the issue of whether or not it would be expedient to include additional tests in the factorial design of the ASVAB. These analyses may point out some broad cognitive areas that the ASVAB does not cover or represent sufficiently. An effort to include tests in the ASVAB that do capture performance in these cognitive domains might improve its predictive validity.

3.2. Participants

Participants were 6751 (1141 female) US Air Force recruits undergoing their sixth week of basic training. The majority of participants (4894) had finished high school (or the equivalent), while 1710 others had some college education, and 147 recruits did not have a high school diploma or GED.[8]

3.3. Design and procedure

Testing was conducted in mixed gender groups of no more than 40. The 46 Kit tests employed in this study were divided into six booklets containing a predetermined mix of seven (and sometimes eight) of these tests. The 10 ASVAB tests made up two more booklets. Participants were administered two booklets, with a break between booklets. All booklet pairs were administered with a target of 200 participants per pair of booklets. Because complete data on the ASVAB were available from the enlistee's records, this information was used instead of the incomplete data obtained during this test session.[9] Using recruits with complete data on the ASVAB and on a pair of the Kit booklets yielded a final sample size of 2897. The 46 Kit tests used in the present study are presented in Appendix B, with the ASVAB tests as described in Appendix A.

3.4. Results and discussion

A series of CFA using missing data methods (Allison, 1987; Muthén, Kaplan, & Hollis, 1987) was performed on the ensuing data set. These CFAs were conducted using the

[8] The reader should bear in mind that because of the matrix sampling procedures employed in Study 2, the effective *N* per pair-wise correlation was approximately 200.

[9] The correlation between tests given at different times was high enough (i.e., consistent with reported test–retest reliabilities) to consider them analogous. Moreover, alternative CFA models' fit to pre- and current ASVAB data did not differ markedly. Time of testing does not appear, therefore, to confound the results reported herein.

Table 3
Standardized loadings for Stratum Ia factors derived from a CFA of Study 2 data

Psychometric tests	Quantitative	Vocabulary	Spatial	Visual	TK	Residual variance
Kit cognitive areas						
General Reasoning	0.98					0.05
Numerical	0.16					0.40
Verbal		0.92				0.15
Flexibility of Closure			1.00			
Spatial			1.00			
Spatial Scanning			0.82			0.10
Visualization				1.00		
Figural Flexibility				1.00		
ASVAB						
General Science					0.31	0.40
Word Knowledge		0.86				0.26
Math Knowledge	0.77					0.41
Arithmetic Reasoning	0.82					0.34
Mechanical Comprehension	0.35				0.58	0.41
Electronic Information					0.65	0.42
Autoshop Information					0.80	0.36

(1) Tests in bold appear in the structure twice. (2) Each kit cognitive area is a Stratum I factor consisting of two tests. These tests are listed in Appendix B.

STREAMS shell (Gustafsson & Stahl, 1996), which interacts with LISREL 8.12 (Joreskog & Sorbom, 1993, used in this study), as well as EQS (Bentler, 1993). All analyses were performed on number-correct scores from the 46 tests comprising the Kit Battery. The CFAs involved testing a number of models. The first was a simple model with 23 correlated factors (corresponding to the 23 primary abilities that the 46 Kit tests used in this study were designed to assess). Next, a hierarchical structure was tested that corresponded to the second-order factors (or domains) hypothesized by Carroll (1993), with the same 23 first-order factors as in the previous analysis and two factors falling in between Strata I and II. The final model (which is the one presented here) reproduced this structure with the 46 Kit scores augmented by the 10 ASVAB test scores.

The factorial structure suggested by Gf/Gc theory, along with previous results from analyses of the ASVAB, and findings for the 23 correlated factors of the Kit were informative in guiding our CFA. In the end, a model having six Stratum II, 33 Stratum I, and five Stratum Ia factors was posited.[10] This model fits very well, given the large sample size with a χ^2 for model = 4352.00, *df* = 1438, and a root mean square error of approximation (RMSEA) = 0.0265. The latter fit statistics incorporates information from the

[10] Carroll (1993, Chap. 15) discusses his theoretical cognitive structure using the Strata I and II concepts. He also discusses conditions under which lower-order factors represent abilities in ‘some sort of limbo between Strata I and II’ (p. 596). For ease of exposition, the factors identified in our study that exhibit this quality are referred to as lying on Stratum Ia.

fit, the sample size, and the *df*, and has a target value of 0.05 (Browne & Cudeck, 1993). This model may even overfit the data, as some researchers have suggested that an RMSEA below 0.03 is indicative of overfit (Gustafsson & Stahl, 1996; Joreskog & Sorbom, 1993). Results from this CFA are presented in Tables 3 (Stratum Ia) and 4 (Stratum II), and a

Table 4
Standardized loadings for Stratum II factors derived from a CFA of Study 2 data

Psychometric constructs	Gf	Gc	SAR	Gv	TSR	Gs	Residual variance
Kit cognitive areas							
Induction	0.99						0.01
Integrative Processes	0.90						0.18
Logical Reasoning	0.87						0.25
Quantitative	0.87						0.24
Memory Span	0.49						0.76
Vocabulary		0.95					0.09
Associational Fluency		0.30			0.66		0.31
Word Fluency		0.18			0.84		0.14
Associational Memory			1.00				
Visual Memory			0.60	0.53			0.24
Visualization				1.00			
Spatial				0.89			0.21
TK				0.51			0.74
Speed of Closure				0.61			0.64
Expressional Fluency					0.84		0.29
Ideational Fluency					0.77		0.41
Flexibility of Use					0.76		0.42
Verbal Closure					0.65		0.58
Figural Fluency					0.28		0.92
Perceptual Speed						0.76	0.42
Numerical						0.71	0.40
Spatial Scanning						0.42	0.10
ASVAB tests							
Coding Speed						0.71	0.50
Numerical Operations						0.74	0.45
General Science		0.65					0.40
Paragraph Comprehension		0.57				0.23	0.64
Electronic Information		0.29					0.42

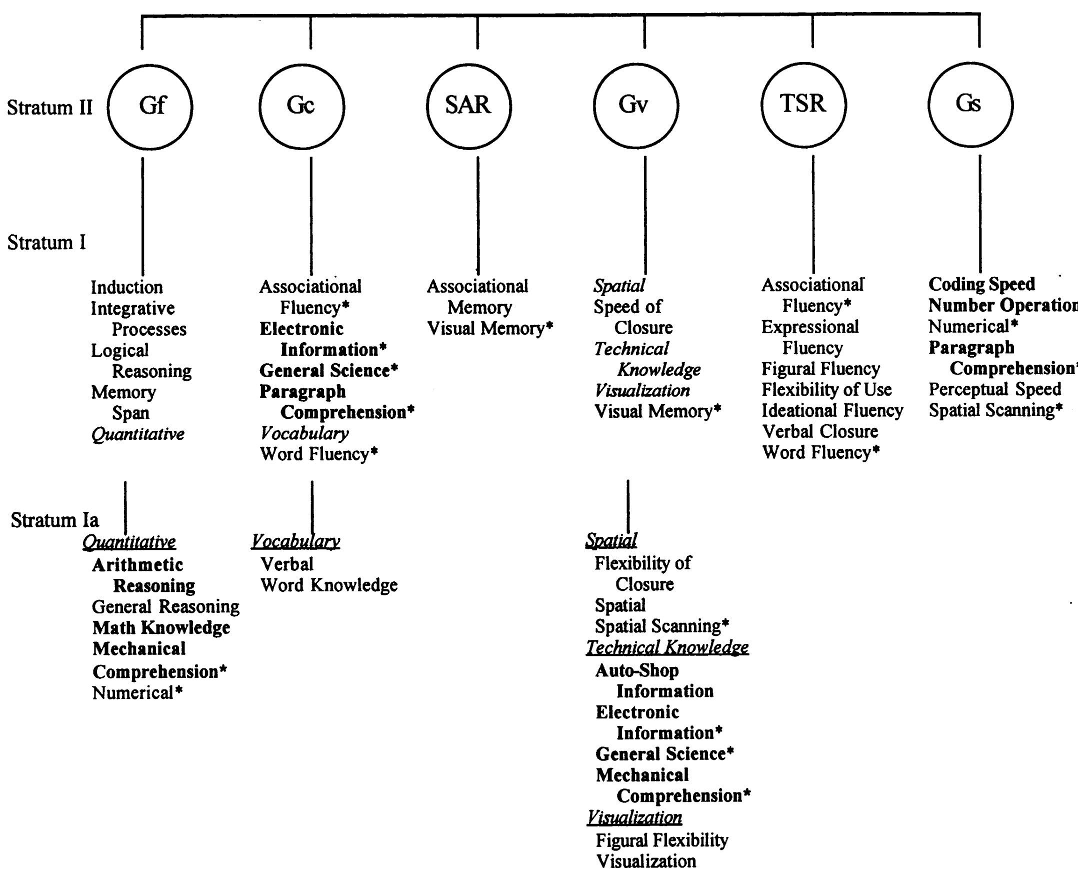
Correlations among Stratum II factors

Factor	Gf	Gc	SAR	Gv	TSR
Gc	.57	–			
SAR	.39	.10	–		
Gv	.74	.35	.19	–	
TSR	.62	.38	.36	.36	–
Gs	.41	– .04	.26	.09	.46

(1) Tests in bold appear in the structure twice. (2) Italicized titles represent Stratum Ia factors. (3) Each Kit cognitive domain is a Stratum I factor consisting of two tests. These tests are listed in Appendix B.

visual representation of the factor structure is depicted in Fig. 2. In both of the tables, a missing residual variance implies that the residual variance parameter was fixed at zero in the analysis.

Examination of Tables 3 and 4 provides supporting evidence for some of the constructs encapsulated by Gf/Gc theory. Thus, one may find, among the factors defined largely by Kit tests, a construct that is clearly fluid intelligence. However, only two ASVAB tests (i.e., Math Knowledge and Arithmetic Reasoning) share any loadings on this factor (and then 'indirectly' through the Quantitative Stratum Ia factor). These



Notes:
(1) Tests with an asterisk (e.g., Numerical*) appear in the structure twice.
(2) Tests in bold (e.g., **Coding Speed**) come from the ASVAB while italicized titles (e.g., *Vocabulary*) represent Stratum Ia factors.
(3) All Stratum II factors are correlated to some (occasionally small) extent (see Table 2). These correlations are indicated by a top bar, which links each factor to one another.

Fig. 2. The factor structure of the ASVAB and Kit Batteries (for details, see Study 2).

findings replicated the major outcome obtained in Study 1 of the present paper. Also, as in Study 1, no tests from the ASVAB loaded on SAR or, for that matter, the second of the broad constructs related to memory — TSR. Notwithstanding and as expected, Gs shared loadings from both the ASVAB and Kit tests, requiring speed in performing tasks of relatively trivial difficulty.

Paralleling the results presented in Study 1, the two remaining factors account for individual differences in performing tests from the ASVAB. One should note that the Gc factor of this study, unlike that in Study 1, fails to share loadings from any of the 'Quantitative' tests found in the ASVAB or, for that matter, the Kit Battery. This supports Carroll's (1993, p. 599) proposition that crystallized intelligence markers are mostly verbal tests, and that when quantitative tests load on this factor, it is because the tests are presented with high verbal processing requirements. Finally, the broad visualization (Gv) factor presented in Table 4 exhibits high loadings from the TK factor of the ASVAB. This outcome is not surprising due to the presence of a large number of tests from the Kit Battery that are markers for the Gv factor. Indeed, loadings from the ASVAB tests testify to the broadness of the visualization factor.

The Stratum II factor intercorrelations are somewhat different from those found in Study 1. The correlation between Gf and Gc is moderate, while the correlations between Gf and both Gv and TSR are relatively high. Indeed, when coefficients presented in Table 4 are rank-ordered, Gf is found, almost without exception, to demonstrate the highest correlation with each of the second-stratum constructs. Elsewhere, the magnitude of correlation between fluid intelligence and sensory and memory processes has been predicated upon the fact that these are all vulnerable to the influences of aging and neurological degradation (Horn & Hofer, 1992). A ready correspondence with this literature is thus compelling. Notwithstanding, the correlation that Gc shares with almost all factors is rather low, indicating that it is likely not as an important component of psychometric *g* as is sometimes envisaged. In lending support to this proposition, it is noteworthy that Gs correlates negatively with this broad acculturated learning factor (see Roberts & Stankov, 1999, who have observed a similar outcome). In sum, the results from Study 2 indicate still further that fluid and crystallized intelligence are structurally independent, and perhaps more importantly that the ASVAB under-represents several broad factors considered crucial in extant models of human cognitive ability.

4. General discussion

Findings from these two studies support the theoretical framework of cognitive ability structure that includes Gf, Gc, and other broad memory and perceptual factors as distinct (but correlated) second-order constructs (see Carroll, 1993; Horn, 1998; Stankov et al., 1995, 2001). This outcome is worth noting in light of the fact that the 'hybrid' model that Carroll (1993) proposed still contained a number of inferences as to the manner in which many cognitive abilities were interrelated and arranged. Carroll (1993) was forced to make this series of inferences because of the practical limitations imposed upon factorial studies requiring both a large database (i.e., number of tests) and appropriate sample size. In support of this proposition, consider the following. The sample sizes that Carroll (1993) had available to him were modest (Median = 198, Table 4.3, p. 118) as were the number of variables employed (Median = 19.6, Table 4.7, p. 123), with studies providing coverage of two (or more) second-stratum constructs appearing infrequently. Further, Carroll's (1993) model derives exclusively from exploratory factor analytic techniques. In light of these limiting features, the degree to which Study 2 data attest to second-stratum constructs is compelling. Thus, the number of tests and participants we were able to examine (using missing data methods) was particularly large, while the CFA solution that we reported reproduced a structure previously founded on exploratory techniques.

It might be objected that failing to model a third-stratum psychometric *g* within the present series of studies constitutes an oversight on our part. However, we contend that the status of this construct is more equivocal than has often times been acknowledged in the recent literature (see Horn, 1998; Pallier, et al., 1999; Roberts et al., 1999, who all make a similar point). Moreover, the two studies were designed to elucidate the nature of second-stratum constructs and the place of the ASVAB within this structure. Even so, results presented in our second study indicate the *g* construct to correspond most closely to the second-stratum fluid intelligence factor.[11] This proposition needs to be contrasted with the notion that some commentators entertain, wherein it is argued 'verbal math is frequently considered the avatar of *g*' (e.g., Stauffer et al., 1996, p. 199; see also Matarazzo, 1972). In light of our findings, it remains an open empirical question whether various claims surrounding the predictive properties of psychometric *g* (in a wide variety of selection contexts) are supported by data obtained from the ASVAB.[12]

In light of the preceding arguments, the present data, perhaps most importantly, call into question the conclusions reached in *The Bell Curve* (cf. Chabris, 1998). In this book, almost an entire chapter is devoted to discussion of the ASVAB, largely because it is on the basis of data collected with this instrument (for the 1980 'Profile of American Youth') that pivotal empirical analyses were conducted (see Herrnstein & Murray, 1994, Appendices 2 and 3, pp. 569–592). In sampling a limited universe of cognitive abilities, which reflect a general acculturated learning factor (rather than psychometric *g* or Gf), the whole of *The Bell Curve* exercise is rendered problematic. The differential crystallized intelligence of the 'underclass' has never been in dispute, and it is the failure of intervention strategies for an ability that is highly malleable, which should primarily have been examined more fully. Moreover, the so-called Flynn effect represents empirical confirmation that fluid abilities increase over time, a point on which the authors of *The Bell Curve* might have remained silent, since this crucial

---

[11] Further analyses of the data in our second study were performed to test the veracity of this claim. These results show that the *g* construct corresponds (near unity) with the second-stratum Gf factor, and that it correlates highly with Gv and TSR factors, yet only moderately with crystallized intelligence. Moreover, the extracted general factor has low loadings on the Gs and SAR constructs.

[12] The problem of factorial invariance also appears in comparing the general factors that might have been obtained from the two studies (e.g., Horn, 1998). Factor intercorrelations presented in Study 1 are indicative of a substantially weaker general factor, with the pattern of loadings notably different from loadings presented in Study 2.

ability was not assessed by Herrnstein and Murray (1994). Further still, the present findings question the types of analysis conducted by Herrnstein and Murray (1994). In particular, using regression analyses with educational attainment and ASVAB scores as predictors of various social criteria, one may reach a conclusion that intelligence (i.e., ASVAB rating) is a superior predictor. This may appear surprising, because it is generally assumed that education should incorporate intelligence, not the other way around. However, if intelligence is defined in terms of Gc, precisely the opposite may be expected, as happened in *The Bell Curve* (see Stankov, 1995).

In a somewhat different vein, it has been suggested that 'processing-oriented' tasks should replace traditional cognitive ability assessment some time in the future (e.g., Kyllonen, 1994). Certain frustrations currently expressed with this undertaking may stem from the factorial composition of tests, such as the ASVAB, with which these measures have so far been analyzed and compared (see Goff, Sawin, & Earles, 1997; Gustafsson & Muthén, 1994). As Ackerman (1996) has recently argued in his PPIK model (intelligence as process, personality, interests and knowledge), there would appear two main types of intelligence: intelligence as process (akin to Gf) and intelligence as knowledge (akin to Gc). It seems plausible that a true test of the efficacy of processing measures awaits more theory-based models of psychometric assessment and certainly ones that take into account the various cognitive strata included in Carroll's (1993) taxonomic model.

Finally, in consideration of its application in personnel selection, the data indicate the ASVAB (and probably other selection tests) is in need of refinement. Revisions to the ASVAB should start at the fundamental level of deciding which cognitive domains to cover. Clearly, some of these domains are more pertinent to the primary aim of the ASVAB (that of predicting performance by enlisted personnel) than others. While tests reflecting crystallized intelligence should be retained in any revision to the battery because they have historically helped predict performance in training schools, there are several constructs that remain poorly operationalized.[13] In particular, two or three prototypical measures of Gf should be included in a revised battery, since current ASVAB measures that load on Gf over-represent the quantitative domain. Indeed, purely on the grounds of practical utility, a strong case may be made for including assessment of all but two second-stratum factors found in Carroll's (1993) taxonomic model.[14] Of these second-order constructs, TSR and broad clerical-perceptual speed (Gs) alone do not seem to fit readily into the present selection requirements of the military. In considering the evolution of changing social demands on cognitive abilities *per se*, it should not pass unnoticed that Gs is currently assessed by two ASVAB subtests — both of which assess a type of performance rendered relatively obsolete by computer technology.

[13] We hasten to add that in order to remain fair to minority groups, crystallized intelligence tests require more careful norming than is generally needed for any other test assessing second-stratum constructs.

[14] Given the importance of oral communication and a likely increase in the use of computerized speech generation and perception, it may especially be useful to supplement the current ASVAB format with measures of speech perception.

Acknowledgments

We would like to thank Professor John B. Carroll for thoughtful comments on an earlier draft of this manuscript.

Appendix A. A brief description of each of the 10 tests comprising the ASVAB follows

1. *General Science*. This test consisted of 25 science-fact items. For example: "Which of the following foods contain the most iron? (a) eggs, (b) liver, (c) candy, or (d) cucumber."
2. *Arithmetic Reasoning*. This test consisted of 30 arithmetic word problems. For example: "Pat put in a total of 16 h on a job during 5 days of the past week. How long is Pat's average workday: (a) 3 h; (b) 3 h; 15 min; (c) 3 h, 18 min; or (d) 3 h, 25 min."
3. *Word Knowledge*. This test contained 35 standard vocabulary items, such as "The wind is *variable* today. (a) mild, (b) steady, (c) shifting, or (d) chilling."
4. *Paragraph Comprehension*. In this test, participants were presented with 15 paragraphs, each one to three sentences long, followed by a multiple-choice response question about the paragraph's content.
5. *Numerical Operations*. This was a 10-min speeded test. The participants task is to respond to 50 (simple) number-fact items (e.g., $2 \times 6 = ?$ (a) 4, (b) 8, (c) 3, or (d) 12).
6. *Coding Speed*. This is another 10-min speeded test. An item consisted of a word followed by five four-digit number strings (e.g., "green. (a) 6456, (b) 7150, (c) 8385, (d) 8930, (e) 9645"). The participant's task was (1) to look up the word's number code in a key consisting of 10 word-code pairs placed at the top of the page, and then (2) select the letter associated with that number code. Coding speed consisted of 84 such items.
7. *Auto and Shop Information (Autoshop)*. This test consisted of 25 questions about automobiles, shop practices, and the conventional use of mechanical tools.
8. *Mathematics Knowledge*. This test consisted of 25 mathematical problems (primarily algebra but also questions concerning area, square roots, percentages, and simple geometry). For example: "If $3X = -5$, then $X = ?$ (a) -2, (b) $-5/3$, (c) $-3/5$, or (d) 3/5."
9. *Mechanical Comprehension*. This test was comprised of 25 questions, normally accompanied by drawings. These questions relate to general mechanical and physical principles.
10. *Electrical Information*. This test contained 20 questions that relate to electrical, radio, and electronics information.

Appendix B. Table showing primary mental ability and corresponding tests assessing each construct (see Study 2)

Descriptions of each primary mental ability may be found in Ekstrom et al. (1976), and also various other sundry publications in which these constructs are discussed (e.g., Carroll, 1993; Horn, 1998). In the latter instances, slightly different nomenclature is sometimes used, although the constructs remain remarkably similar across major theoretical instantiations.

Cognitive area	ETS Kit Test
Associational Fluency	Controlled Associations
	Opposites
Associative Memory	Object-Number
	Picture-Number
Expressional Fluency	Arranging Words
	Making Sentences
Figural Flexibility	Storage
	Toothpicks
Figural Fluency	Elaboration
	Ornamentation
Flexibility of Closure	Copying
	Hidden Patterns
Flexibility of Use	Combining Objects
	Making Groups
General Reasoning	Arithmetic Aptitude
	Necessary Arithmetic Operations
Ideational Fluency	Thing Categories
	Topics
Induction	Figure Classification
	Letter Sets
Integrative Process	Calendar
	Following Directions
Logical Reasoning	Diagramming Relationships
	Nonsense Syllogisms
Memory Span	Auditory Letter Span
	Auditory Number Span
Number	Addition
	Subtraction and Multiplication
Perceptual Speed	Finding A's
	Number Comparison
Spatial Orientation	Card Rotations
	Cube Comparisons
Spatial Scanning	Map Planning
	Maze Tracing Speed
Speed of Closure	Concealed Words
	Gestalt Completion
Verbal Closure	Incomplete Words
	Scrambled Words

Verbal Comprehension	Vocabulary I
	Vocabulary II
Visualization	Paper Folding
	Surface Development
Visual Memory	Building Memory
	Map Memory
Word Fluency	Word Beginnings
	Word Endings

References

Ackerman, P. L. (1996). A theory of adult intellectual development: process, personality, interests, and knowledge. *Intelligence*, *22*, 227–257.

Allison, P. D. (1987). Estimation of linear models with incomplete data. In: C. Clogg (Ed.), *Sociological methodology 1987* (pp. 71–103). San Francisco: Jossey Bass.

Anastasi, A. (1988). *Psychological testing* (6th ed.). New York: Macmillan.

Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). NJ: Prentice-Hall.

Bentler, P. M. (1993). *EQS: structural equations program manual*. Los Angeles: BMDP Statistical Software.

Binet, A., & Simon, T. (1916/1983). *The development of intelligence in children*. Salem, NH: Clyer.

Boring, E. G. (1923). Intelligence as the tests test it. *The New Republic*, *35*, 35–37 (6/6/23).

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In: K. A. Bollen, & J. S. Long (Eds.), *Testing structural equation models* (pp. 111–123). Newbury Park, CA: Sage Publication.

Carroll, J. B. (1993). *Human cognitive abilities: a survey of factor-analytic studies*. New York: Cambridge Univ. Press.

Carroll, J. B. (1997). Psychometrics, intelligence and public perception. *Intelligence*, *24*, 25–52.

Chabris, C. F. (1998). IQ since "The Bell Curve". *Commentary*, 33–40 (August).

Cronbach, L. J. (1984). *Essentials of psychological testing* (4th ed.). New York: Harper & Row.

Daniel, M. H. (1997). Intelligence testing: status and trends. *American Psychologist*, *52*, 1038–1045.

Davies, M., Stankov, L., & Roberts, R. D. (1998). Emotional intelligence: in search of an elusive construct. *Journal of Personality and Social Psychology*, *75*, 989–1015.

Ekstrom, R. B., French, J. W., Harmon, H. H., & Derman, D. (1976). *ETS kit of factor-referenced cognitive tests*. Princeton, NJ: Educational Testing Service.

Frank, G. (1983). *The Wechsler enterprise: an assessment of the development, structure, and use of the Wechsler tests of intelligence*. New York: Pergamon.

Goff, G. N., Sawin, L., & Earles, J. (1997). *The factor structure of two abilities tests: APT vs. ASVAB*. Paper presented at the 105th Annual American Psychology Association Conference, Chicago, IL, August.

Gregory, R. J. (1996). *Psychological testing: history, principles, and applications* (2nd ed.). Boston, MA: Allyn & Bacon.

Gustafsson, J.-E. (1984). A unifying model for the structure of intellectual abilities. *Intelligence*, *8*, 179–203.

Gustafsson, J.-E. (1988). Hierarchical models of individual differences in cognitive abilities. In: R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence,* vol. IV (pp. 35–71). Hillsdale, NJ: Erlbaum.

Gustafsson, J.-E., & Muthén, B. O. (1994). *The nature of the general factor in hierarchical models of the structure of cognitive abilities: alternative models tested on data from regular and experimental military enlistment tests*. Technical Report: University of Göteborg, Sweden.

Gustafsson, J.-E., & Stahl, P.-A. (1996). *STREAMS user's guide: version 1.6 for Windows*. Molndal, Sweden: MultivariateWare.

Hearnshaw, L. S. (1951). Exploring the intellect. *British Journal of Psychology*, *42*, 315–321.

Herrnstein, R. J., & Murray, C. (1994). *The Bell Curve: intelligence and class structure in American life*. New York: Free Press.

Horn, J. L. (1985). Remodeling old models of intelligence. In: B. B. Wolman (Ed.), *Handbook of intelligence: theories, measurements and applications* (pp. 267–300). New York: Wiley.

Horn, J. L. (1998). A basis for research on age differences in cognitive capabilities. In: J. J. McArdle, & R. W. Woodcock (Eds.), *Human cognitive abilities in theory and practice* (pp. 57–91). Mahwah, NJ: Erlbaum.

Horn, J. L., & Cattell, R. B. (1967). Age differences in fluid and crystallized intelligence. *Acta Psychologica*, *26,* 107–129.

Horn, J. L., & Hofer, S. M. (1992). Major abilities and development in the adult period. In: R. J. Sternberg, & C. Berg (Eds.), *Intellectual development* (pp. 44–99). New York: Cambridge Univ. Press.

Horn, J. L., & Noll, J. (1994). A system for understanding cognitive capabilities: a theory and the evidence on which it is based. In: D. K. Detterman (Ed.), *Current topics in human intelligence,* vol. IV (pp. 151–203). Norwood, NJ: Ablex.

Humphreys, L. G. (1962). The organization of human abilities. *American Psychologist*, *17*, 475–483.

Hunt, E. (1995). *Will we be smart enough? A cognitive analysis of the coming workforce*. New York: Russell Sage Foundation.

Jensen, A. R. (1979). *g*: outmoded theory or unconquered frontier? *Creative Science and Technology*, *2*, 16–29.

Jensen, A. R. (1985). Armed services vocational aptitude battery. *Measurement and Evaluation in Counseling and Development*, *18*, 32–37.

Joreskog, K. G., & Sorbom, D. (1993). *LISREL 8: user's reference guide*. Chicago: Scientific Software International.

Kaplan, R. M., & Saccuzzo, D. P. (1997). *Psychological testing: principles, applications, and issues* (4th ed.). Pacific Grove, CA: Brooks-Cole.

Kaufman, A. S. (1979). *Intelligent testing with the WISC-R*. New York: Wiley.

Kaufman, A. S., & Kaufman, N. L. (1993). *Kaufman adolescent and adult intelligence test*. Circle pines, MN: American Guidance.

Kyllonen, P. C. (1994). CAM: a theoretical framework for cognitive abilities measurement. In: D. K. Detterman (Ed.), *Current topics in human intelligence,* vol. IV (pp. 307–359). Norwood, NJ: Ablex.

Marshalek, B., Lohman, D. F., & Snow, R. E. (1983). The complexity continuum in the radex and hierarchical models of intelligence. *Intelligence*, *7*, 107–127.

Matarazzo, J. D. (1972). *Wechsler's measurement and appraisal of adult intelligence* (5th ed.). Baltimore: Williams & Wilkins.

Matarazzo, J. D. (1992). Psychological testing and assessment in the 21st century. *American Psychologist*, *47*, 1007–1018.

McArdle, J. J., & Horn, J. L. (1983). *Validation by systems modeling of WAIS abilities*. Baltimore: National Institute of Aging.

McGrew, K. S., & Flanagan, D. P. (1998). *The intelligence test desk reference (ITDR): Gf-Gc cross-battery assessment*. Boston, MA: Allyn & Bacon.

Murphy, K. (1984). Armed Services vocational aptitude battery. In: D. J. Keyser, & R. C. Sweetland (Eds.), *Test critiques,* vol. 1. Kansas City, MO: Test Corporation of America.

Murphy, K. R., & Davidshofer, C. O. (1998). *Psychological testing: principles and applications* (4th ed.). Upper Saddle River, NJ: Prentice-Hall.

Muthén, B. O., Kaplan, D., & Hollis, M. (1987). On structural equation modeling with data that are not missing completely at random. *Psychometrika*, *52*, 431–462.

Naglieri, J. A. (1997). IQ: known and unknowns, hits and misses. *American Psychologist*, *52*, 75–76.

Naglieri, J. A., & Das, J. P. (1997). *Cognitive assessment system*. Chicago: Riverside Publishers.

Neisser, U. (1997). Never a dull moment. *American Psychologist*, *52*, 79–81.

Neisser, U., Boodoo, G., Bouchard, T. J. Jr., Boykin, A. W., Brody, N., Ceci, S. J., Halpern, D. F., Loehlin, J. C., Perloff, R., Sternberg, R. J., & Urbina, S. (1996). Intelligence: knowns and unknowns. *American Psychologist*, *51*, 77–101.

Pallier, G., Roberts, R. D., & Stankov, L. (2001). Biological versus psychometric intelligence: Halstead's (1947) distinction re-visited. *Archives of Clinical Neuropsychology*, *13*, 205–226.

Ree, M. J., & Carretta, T. R. (1994). Factor analysis of ASVAB: confirming a Vernon-like structure. *Educational and Psychological Measurement*, *54*, 457–461.

Ree, M. J., & Carretta, T. R. (1995). Group differences in aptitude factor structure on the ASVAB. *Educational and Psychological Measurement*, *55*, 268–277.

Ree, M. J., & Carretta, T. R. (1996). The central role of *g* in military pilot selection. *The International Journal of Aviation Psychology*, *6*, 11–123.

Ree, M. J., & Earles, J. A. (1992). Intelligence is the best predictor of job performance. *Current Directions in Psychological Science*, *1*, 86–89.

Ree, M. J., & Earles, J. A. (1993). *g* is to psychology what carbon is to chemistry: a reply to Sternberg and Wagner, McClelland and Calfee. *Current Directions in Psychological Science*, *2*, 11–12.

Ree, M. J., Earles, J. A., & Teachout, M. S. (1994). Predicting job performance: not much more than *g*. *Journal of Applied Psychology*, *79*, 518–524.

Roberts, R. D. (1997). Fitts' law, movement time, and intelligence. *Personality and Individual Differences*, *23*, 227–246.

Roberts, R. D., Pallier, G., & Goff, G. N. (1999). Sensory processes within the structure of human cognitive abilities. In: P. L. Ackerman, P. C. Kyllonen, & R. D. Roberts (Eds.), *Learning and individual differences: process, trait, and content determinants* (pp. 339–370). Washington, DC: American Psychological Association.

Roberts, R. D., & Stankov, L. (1999). Individual differences in speed of mental processing and human cognitive abilities: towards a taxonomic model. *Learning and Individual Differences*, *11*, 1–120.

Roberts, R. D., Stankov, L., Pallier, G., & Dolph, B. (1997). Charting the cognitive sphere: tactile and kinesthetic performance within the structure of intelligence. *Intelligence*, *25*, 111–148.

Senior, G. (1996). *Recent developments in WAIS-R and WMS-R interpretation*. Paper presented at the 31st Annual Conference of the Australian Psychological Society, Sydney, NSW, Australia, September.

Stankov, L. (1988). Aging, attention, and intelligence. *Psychology and Aging*, *3*, 59–74.

Stankov, L. (1995). LaBELLe curve turned ugly: messages from Herrnstein and Murray's book "The Bell Curve". *Australian Psychologist*, *30*, 218–220.

Stankov, L., Boyle, G. J., & Cattell, R. B. (1995). Models and paradigms in intelligence research. In: D. Saklofske, & M. Zeidner (Eds.), *International handbook of personality and intelligence* (pp. 15–43). New York: Plenum.

Stankov, L., Seizova-Cajic, T., & Roberts, R. D. (2001). Tactile and kinesthetic perceptual processes within the taxonomy of human cognitive abilities. *Intelligence*, *28*, 1–29.

Stauffer, J. M., Ree, M. J., & Carretta, T. R. (1996). Cognitive-components tests are not much more than *g*: an extension of Kyllonen's analyses. *The Journal of General Psychology*, *123*, 193–205.

Sterling, M., Goff, G. N., & Sawin, L. L. (1997). *Assembling objects: addition to the ASVAB?* (Unpublished manuscript).

Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1985). *Stanford–Binet intelligence scale* (4th ed.). Chicago: Riverside Publishing.

U.S. Department of Defense. (1982). *Profile of American youth: 1980 nationwide administration of the Armed Services Vocational Aptitude Battery*. Washington, DC: Office of the Assistant Secretary of Defense (Manpower, Reserve Affairs, and Logistics).

U.S. Department of Defense. (1984). *Test manual for the Armed Services Vocational Aptitude Battery (DoD 1340.12AA)*. North Chicago, IL: U.S. Military Entrance Processing Command.

Vernon, P. E. (1960). *The structure of human abilities*. London: Methuen (Revised edition).

Wechsler, D. (1981). *Manual for the Wechsler adult intelligence scale-revised*. New York: The Psychological Corporation.

Woodcock, R. W., & Johnson, M. B. (1989). *Woodcock–Johnson tests of cognitive ability: standard and supplemental batteries*. Chicago: Riverside Publishers.

Wothke, W., Bock, R. D., Curran, L. T., Fairbank, B., Augustin, J. W., Gillet, A. H., & Guerrero, C. Jr. (1991). *Factor analytic examination of the armed services vocational aptitude battery (ASVAB) and the kit of factor-referenced tests*. Brooks AFB: Air Force Human Resources Laboratory, Manpower and Personnel Division (AFHRL-TR-90-67).

Yerkes, R. M. (Ed.) (1921). *Psychological examining in the United States Army: Memoirs of the National Academy of Sciences, 15*. Washington D.C.: The National Academy of Science.