
Pergamon

Learning and Individual Differences
12 (2000) 105–121

Learning and Individual Differences

Explorations of crystallized intelligence Completion tests, cloze tests, and knowledge

Phillip L. Ackerman*, Margaret E. Beier, Kristy R. Bowen

School of Psychology, Georgia Institute of Technology, Room 130, Psychology Building, 274 5th Street, MC 0170, Atlanta, GA 30332-0170, USA

Abstract

An attempt is made to reconcile two historically important tools for the assessment of intelligence and the prediction of academic achievement with extant theories of verbal–crystallized–knowledge aspects of adult abilities. A study of 167 adults ranging in age from 18 to 69 reasserts the importance of individual differences in completion test and cloze test performance in accounting for both measures of crystallized intelligence (Gc) and four scales of knowledge (biology, U.S. history, U.S. literature, and technology). The completion tests were found to account for all of the variance in Gc and knowledge that the cloze tests accounted for, and resulted in incremental predictive validity for both domains. In addition, completion and cloze tests were found to have a suppressor effect on the relationship between Gc and Age. We note that C. Spearman's [*The nature of "Intelligence" and the principles of cognition.* New York: MacMillan (1927).] assertion, namely that the completion test had higher correlations with intelligence than any other measure. Our results suggest that abstract reasoning may be far less useful in predicting learning and performance than the completion test is.

Keywords: Intelligence; Cloze; Crystallized ability; Knowledge; Aging

1. Background

One of the things we can learn from a study of the history of differential psychology over the past 100 years is that even before the passage of the first 25 years of the 20th century, a great amount of theory and experimentation had been devoted to the assessment of

* Corresponding author. Fax: +1-404-894-6904.
E-mail address: phillip.ackerman@psych.gatech.edu (P.L. Ackerman).

1041-6080/00/$ – see front matter
PII: S1041-6080(00)00034-0

intelligence and to the development of measures for predicting learning and academic success. It is interesting to note how much of this valuable work has been forgotten or ignored in modern theory and empirical research. In the current paper, we focus on a measure that was first developed and administered by Ebbinghaus (1896–1897) for predicting school performance, used extensively in the early 1900s, and then lost until a similar measure was developed by Taylor (1953) for assessing readability and later used for ability assessment. The goal of this investigation is to reconsider these two tests within extant theories of adult intelligence in general, and crystallized intelligence (Gc) in particular. A short review of the history of these measures below will precede a presentation of the current study.

1.1. The completion test

In his classic study of fatigue on school performance, Ebbinghaus (1896–1897) created and administered a *Combinationsgabe* test that has variously been translated as the completion test of the combination test (Carroll, 1982; Whipple, 1921). According to Wilhelm's translation of Ebbinghaus "pupils get prose text with gaps (e.g., missing syllables, missing parts of syllables, and missing words) and are asked to fill in gaps in a meaningful way." The number of syllables and parts of syllables that need to be filled in is indicated for each gap. Pupils worked on each text section for 5 min. The texts were from Swift's Gulliver and from Nettlebeck's *Citizens of Colberg*. Terman's (1906) adaptation of the test, which he initially called the Mutilated Text, involved two different instantiations. The first was a story with missing words, which the examinees were asked to complete as they thought it ought to be done (p. 342). In the second instantiation of the test, Terman stated: "In order to rob the test of its puzzle nature, a second trial was given ... but in this case the complete story was read aloud to the subject first. He therefore knew the general sense and had a much narrower field to hunt over in the search for suitable words" (Terman, 1906, p. 344). In Terman's study of two dozen boys among the brightest or most stupid that could be found in the public schools within easy distance of Clark University, (p. 314) he determined that: The results of the two tests run closely parallel. With certain exceptions the two groups are widely separated in each (p. 347).

Terman reported that "Ebbinghaus regards this test as a reliable measure of intellectual ability. ... Intellectual ability consists in the elaboration of the whole into its worth and meaning by means of many-sided combination, correct, and completion of numerous kindred associations.... It is a *combination* activity... I am inclined to think that somewhat mechanical activities like memory and association, as distinguished from synthetic or combinative processes, play a relatively more important role in this test than Ebbinghaus assigns to them. Indeed, verbal memory, in the broad sense, would seem to be the chief factor in success. It gives what we term *fluency* in language. Verbal memory also means ability to carry the story as a whole, and therefore to see the connections in meaning running through it" (Terman, 1906, p. 347).

Early use of the completion test (either with prior presentation of the text or without) suggested that the test was an excellent indicator of general intelligence. For example, Krueger and Spearman (1907) reported that the completion test had a corrected *g*-loading of .97 — other substantial correlations were reported for measures of memory, opposites,

adding, and overall academic success (see Whipple, 1921). In the basic form of the completion test, using printed text and written responses — it was not ultimately suitable for one-on-one tests that de-emphasized reading and writing, such as the Binet–Simon method. Interestingly, though, Spearman (1927, pp. 3–4) reported that the completion test, even though it produced, as a single test, higher correlations with intelligence than any other measure ... its use as a single test has since been discounted by the fact that the whole employment of such tests singly has everywhere been abandoned. Given the strong associations with other abilities and with school performance, it is at present a puzzle why continued use of the completion test in group testing of intelligence did not continue. (This is an object of current historical investigation [Ackerman, Beier, Wittmann, & Wilhelm, in preparation.].)

1.2. The cloze test

Without any apparent reference to the earlier completion test, and from a completely different perspective, Taylor (1953, 1954) introduced the cloze procedure. In the cloze procedure, sample text passages are mutilated (Taylor, 1957, p. 19) by deleting the same number of words from each. The deletion process selects words at random or counts out every *n*th one. The passages are reproduced with some standard size of blank in place of each missing word, and all mutilated materials are administered to all subjects in the test group ... The subjects are asked (a) to guess what the missing words are, and (b) to write their guesses in the corresponding blanks (Taylor, 1957, p. 19). Taylor actually developed the cloze procedure for assessing the readability of text passages — mainly as a contrast to other readability indexes (such as that of Flesch). The inspiration for the procedure was the then-new information theory approach to psychological issues (Taylor, 1954; for a review of the general approach, see for example, Attneave, 1959). That is, if a passage of text is considered as a message to be transmitted from the writer to a receiver (the reader), the quantified uncertainty of the message (or the converse, the message's redundancy) could be computed as the number of words that can be correctly inferred from context in a text that has been mutilated (by deleting words randomly or every *n*th word). In an early study, Taylor (1957) determined that, in addition to providing relative readability indices, the cloze procedure could also be used to index the relative ability of the reader in comparison to other readers in the context of mental ability, prior knowledge, and knowledge after study. According to Taylor ... an individual's cloze performance appears to depend heavily on how well he understands the meaning of the materials administered — hence on the factors that affect comprehension, such as general language facility, specific knowledge, and vocabulary relevant to the materials at hand, native ability to learn, attention, motivation, and so on (Taylor, 1957, p. 19).

1.3. Crystallized intelligence

In the modern history of intelligence theory, there have been several frameworks that have more or less converged on an omnibus factor of intelligence that subsumes experiential–educative–acculturation influences (Horn & Cattell, 1966). This factor was first empirically

realized in early group-factor studies (e.g., Kelly, 1928; Thomson, 1946) as a broad verbal factor, and it is one major component of many broad intellectual ability assessments (e.g., the verbal composite of the Wechsler Adult Intelligence Scale and its revisions; Wechsler, 1958). From a theoretical perspective, this factor was generalized at a higher level as Intelligence B (Hebb, 1942), Gc (Cattell, 1943), and verbal/educational (Vernon, 1950). Prototypical subfactors that load highly on this higher-order intelligence factor include verbal comprehension, lexical knowledge, and reading comprehension (e.g., see Carroll, 1993). For the sake of convenience, we refer to this complex of abilities as Gc.

Among major intelligence theories, those of Cattell (1943) and Hebb (1942) emphasize that the Gc factor of intelligence is well-maintained in adulthood (with positive correlations often found in cross-sectional samples of adults of widely differing ages). The age pattern for Gc of increases with age is often contrasted with factors of abstract reasoning, math abilities, and spatial abilities — which tend to show decreases in adult performance as age increases after the early 20s (e.g., see Cattell, 1971, 1987; Horn & Cattell, 1967).

1.4. Gc and cloze

In Carroll's (1993) seminal reanalysis of a large corpus of factor-analytic research on human abilities, the completion test merited only a brief mention. Carroll stated that the completion test resembled what today would be called a *cloze* test (Carroll, 1993, p. 33). The cloze test received a little more discussion compared to the completion test (pp. 167–168). Six datasets were found that revealed cloze-ability factors, though only one study was reported with examinees over the age of 16 (Weaver & Kingston, 1963). In his reanalysis and reassessment of the literature, Carroll placed the cloze tests in the category of written language and productive skills, (p. 147) associated with the domain of language. Carroll concluded, however, that "There is thus little critical information on the nature of a cloze ability factor" (p. 167).

Given the paucity of other data on the topic, a closer look was warranted for the Weaver and Kingston (1963) study. These authors administered 18 tests (including Carroll's Modern Language Ability Test [MLAT] battery, eight different instantiations of the cloze test, and five other verbal ability tests) to 160 junior-class university students. The authors claimed that the data revealed separate factors of verbal comprehension, cloze, and a factor of MLAT tests they termed rote memory, flexible retrieval (p. 256–259). These authors concluded, on the basis of a factor analysis with a Varimax rotation, that: The cloze tests are related only moderately to the verbal comprehension factor (p. 259), and that Factor III (the MLAT tests) contains most of the variance supplied by the MLAT and little else (p. 260). Of course, historical hindsight (and the use of a modern desktop computer) shows how the independence of the cloze factor was substantially overstated by Weaver and Kingston. In our own reanalysis of the correlations reported in their article, we computed an oblique rotation of the factors, rather than the orthogonal rotation used by Weaver and Kingston. Even though a similar factor pattern was found in our reanalysis (that is the separate cloze factor along with the verbal comprehension and MLAT factors), the correlations among the factors were positive and substantial. The cloze factor correlated r=.412 with the verbal comprehension factor and r=.517 with the MLAT factor. Also, the

verbal comprehension and MLAT factors were substantially correlated, r=.624. All together, these results seem to indicate that the cloze tests were well-placed as a lower-order factor associated with Gc.

1.5. Knowledge

Although theorists such as Cattell (1957) have emphasized the *breadth* of Gc, indicating that Gc includes both broad cultural knowledge and more narrow sources of knowledge (such as occupational and avocational knowledge), few investigators have attempted to develop assessments of knowledge that tap the diverse domains of knowledge that persons presumably develop and maintain over the adult life-span (see Ackerman, 1996 for a review and discussion of this topic). Recently, Ackerman and colleagues have developed and administered a large battery of knowledge tests to adults across an age range from to 18 to 70 years old. In two studies (Ackerman, 2000; Ackerman & Rolfhus, 1999), it was found that knowledge of sciences, humanities, civics, and technology were well-preserved in adult samples — generally concordant with the hypotheses of Cattell (1957) and Hebb (1942). However, these studies also concluded that knowledge is something more than Gc as typically assessed. Specifically, only about 50% of the variance in various knowledge scales was accounted for by measures of Gc. Thus, knowledge represents an important component of adult intellect that is separable from traditional measures of Gc. Given the importance of knowledge in learning and transfer (e.g., see Ferguson, 1956), the separability of knowledge and Gc suggests that both domains should be considered when attempting to predict adult learning and intellectual development (Ackerman, 1998).

2. Study overview

The current study was designed to further explore the complex of abilities that make up the verbal–educational–knowledge component of adult intelligence. In addition to samples of tests that serve as markers for Gc and for knowledge factors, the study aimed to evaluate the techniques of cloze and completion tests as measures that may bridge the domains of Gc and knowledge. On the one hand, various Gc tests call for both recall, recognition, and real-time verbal reasoning. On the other hand, knowledge scales as currently instantiated primarily tap recognition and, to some degree, recall of declarative knowledge.

Based on the historical literature, cloze tests involve "comprehension, such as general language facility, specific knowledge, and vocabulary relevant to the materials at hand" (Taylor, 1957, p. 19). The completion test has had various representations of the underlying determinants (e.g., Ebbinghaus, 1896–1897; Whipple, 1921), ranging from general intelligence to "general command of language, verbal memory, fluency, and so on" (Terman, 1906, p. 347). In our instantiation of the completion tests, though, we sought to make the tests parallel to the cloze procedure — with a single exception, namely, that we read the passage in its entirety to the examinees just prior to the test phase. In doing so, we expected that the completion test would tap the same underlying processes of the cloze tests *and* add a listening and recall component. This procedure would allow us to ascertain whether the

listening and recall component added significantly to the explanation of either individual differences in Gc or knowledge. Moreover, we expected that age effects (in a cross-sectional study) might be highlighted in the completion test scores, given that working memory measures have been shown to substantially decline with age in adult samples (e.g., see Salthouse, 1996).

Thus, our main goal was to examine the relations among Gc, knowledge, and Age variables and individual differences in performance on the cloze and completion tests in a cross-sectional study of adults ranging in age from 18 to 70. Along the way, we aimed to provide descriptive and reliability information regarding these two tests that have infrequently been studied in the context of individual differences in abilities.

3. Method

3.1. Participants

One hundred and sixty-seven participants were recruited through local and community newspapers as well as a student pool at the Georgia Institute of Technology. Inclusion criteria were normal or corrected-to-normal vision, hearing and motor coordination, and all participants were required to have at least some college experience. Participants ranged in age from 18 to 69 years (mean = 37.4 years, S.D. = 13.4 years). The year that participants received their high school diplomas (or the general equivalency diploma) ranged from 1948 to 1999.

3.2. Apparatus

The study included an ability battery and knowledge scales. Demographic data including age, gender, and education were collected through a biographical questionnaire that included other measures that were part of a larger study. The ability battery was administered using paper and pencil tests in a classroom-like setting. Test instructions and start/stop timings were administered with prerecorded minidiscs over a public address system. Knowledge scales were administered with a computer program running under Windows 98 on Pentium computers with standard keyboards. Up to 16 participants were tested at one time.

3.3. Measures

3.3.1. Ability battery

The ability battery included five tests designed to provide assessment of Gc, as follows:

1. Extended range vocabulary (ETS Kit, Ekstrom, French, Harman, & Derman, 1976),
2. information test (Wechsler Adult Intelligence Scales-Revised; Wechsler, 1981),
3. comprehension (Multidimensional Aptitude Battery [MAB]; Jackson, 1985),
4. synonyms (MAB, Jackson, 1985), and
5. word beginnings (Ekstrom et al., 1976).

3.3.2. Cloze tests

Four cloze tests were also included in the ability battery. The four cloze tests were constructed from passages selected from college-level textbooks, as follows:

1. U.S. history (Rice, Krout, & Harris, 1991, p. 25),
2. U.S. literature (DiYanni, 1994, p. 55),
3. biology (Campbell, 1996, p. 913), and
4. narrative (DiYanni, 1994, p. 346).

Passages were originally selected to be around 250 words in length (actual mean = 255.5 words). Following the technique originated by Taylor (1953), a structural (Ohnmacht, Weaver, & Kohler, 1970) cloze test was constructed. This entailed leaving the first and the last sentences of the passage intact. Starting with the second sentence, every fifth word was deleted (regardless of its grammatical or contextual relationship) and replaced with an underlined blank 10 spaces long. This procedure has been shown by Taylor (1957) to be superior to lexical deletion (in which words are deleted based on their relationship to the text) in efficiency of construction as well as equal or superior in its correlation to some ability measures. Cloze tests in this study included mean = 41 blanks. Participants were instructed to read through the passage and fill in the blanks with the words that best fit into the sentence. If participants did not know the exact words that fit in the blank, they were instructed to guess. Participants were given 10 min to complete each cloze test.

Credit was given for either the actual missing word or for words that fit the gist of the paragraph (and were grammatically correct in the context of the text). Scores for the cloze tests reflected the total number of actual or gist words. Participants were not penalized for obvious misspellings of words (i.e., when the misspelling did not change the meaning of the intended word). No credit or penalty was given to words that did not make sense within the context of the paragraph or for words that were grammatically incorrect.

3.3.3. Completion tests

A procedure identical to that of developing cloze tests was used to develop the completion tests.

1. U.S. history (Rice et al., 1991, p. 157),
2. U.S. literature (DiYanni, 1994, p. 77),
3. biology (Campbell, 1996, p. 227), and
4. narrative (DiYanni, 1994, p. 362).

The completion test passages averaged 248.5 words in length with an average of 42 blanks. Completion tests differed from cloze tests in their administration. Specifically, participants were instructed to listen to the passage read in its entirety, without looking at the completion test form. After the passage was read, participants were shown the completion test form and instructed to fill in as many of the missing words as possible. If they did not remember the exact words, participants were instructed to guess. Participants were given 8 min to fill in the completion test form. Scoring for the completion tests was different from that of the cloze

tests, where participants received equivalent credit for exact matches and gist matches. Completion tests were scored by giving participants two points for every exact match and one point for words fitting the gist meaning of the paragraph (as long as the words were grammatically correct). *Thus, one difference between the scoring of the cloze and completion tests was that extra credit was accorded in the completion test for the exact word in the text*. Again, participants were not penalized for obvious misspellings of words.

3.3.4. Biographical questionnaires

The biographical questionnaire collected information regarding participant gender, ethnic background as well as academic history. It also included other measures that were part of another study not reported here.

3.3.5. Knowledge tests

A subset of the knowledge scales used in previous research (Ackerman, 2000; Ackerman & Rolfhus, 1999; Rolfhus & Ackerman, 1999) were used in this study. These knowledge scales were developed mostly with content from the College Level Examination Program (CLEP) and Advanced Placement (AP) tests. Because CLEP/AP tests assess knowledge equivalent to completion of college-level courses, each scale was expanded to include easier items to differentiate participants with less knowledge in these domains than the typical CLEP population. The knowledge scale battery and its derivation are described in greater detail in Rolfhus and Ackerman (1999). The particular tests used in this study were in the domains of:

1. U.S. history. This test covers U.S. history from pre-Revolutionary times to the present (Ackerman & Rolfhus, 1999).
2. U.S. literature. This test covers a broad range of American writers, playwrights, and poets from Revolutionary times to the present. Questions require identification of authors and works from Walt Whitman to Kurt Vonnegut, and an understanding of literary styles and movements such as transcendentalism (Ackerman & Rolfhus, 1999).
3. Biology. This test covers a broad range of biology, at the cellular, organismal, and ecological levels. Questions range from an understanding of food chains to the function of meiosis and mitosis (Ackerman & Rolfhus, 1999).
4. Technology. This test covers a wide range of modern technologies. Questions range from how microwave ovens and televisions work to an understanding of super-conductivity. The test was originally administered in Ackerman and Rolfhus (1999), but was updated and expanded in the current study.

The first three knowledge scales used in this study were designed to parallel the content of three of the completion and cloze tests. They were constructed in a power format where the easiest items were administered first, followed by items in order of increasing difficulty. When a participant incorrectly answered three in a row, the particular test was terminated and another started. The last scale (Technology) involved administration of the entire scale, in randomized order (mainly because new items had been added to the battery, and the study was used to obtain new normative data from all items).

3.4. Procedure

The study involved a single 3 1/2-h session. The session started with 2 1/2 h of ability tests (experimenter-paced), followed by a 5-min break, and then 1 h of computerized knowledge scale administration (participant-paced).

4. Results

4.1. Descriptive statistics and reliabilities

Number of items, means, and standard deviations for the ability tests, and knowledge scales are presented in Table 1. For the four cloze tests, average alternate form reliability was mean $r = .74$ (based on a mean of r-to-z transforms, and a reverse z-to-r transformation of the mean z); for the four completion tests, average alternate form reliability was mean $r = .77$; for the four knowledge scales, the average intertest correlation was mean $r = .54$. At the aggregate

Table 1
Number of items, means, and standard deviations for the Gc tests, cloze tests, completion tests, and knowledge scales

	Total number of items	mean	S.D.
Gc			
1. Vocabulary	48	21.59	10.17
2. Information	24	14.02	4.51
3. Word beginnings	open-ended	24.40	8.53
4. Comprehension	28	20.13	4.21
5. Synonyms	34	24.57	5.77
Cloze tests			
6. Cloze #1 — biology	37	16.16	7.19
7. Cloze #2 — U.S. Literature	39	23.08	7.88
8. Cloze #3 — U.S. history	42	23.07	8.50
9. Cloze #4 — narrative passage	46	33.87	8.63
Completion tests			
10. Completion #1 — biology	44[a]	41.83	18.25
11. Completion #2 — U.S. literature	45[a]	44.93	15.92
12. Completion #3 — U.S. history	37[a]	32.33	13.59
13. Completion #4 — narrative passage	42[a]	54.71	12.18
Knowledge scales			
14. U.S. history	123	44.39	25.80
15. U.S. literature	79	36.56	18.83
16. Biology	73	13.02	10.53
17. Technology	90	43.16	13.11

[a] Maximum score equals (number of items) × 2 (i.e., 2 points were given for each response with exact word).

level, composite scores for Gc, cloze, completion, and knowledge all had internal consistency reliabilities (Cronbach's) exceeding .80.

4.2. Demographic variables

Correlations among all of the separate tests, the composite variables, and demographic variables are presented in Table 2. Concordant with the extant literature, Gc showed a positive significant correlation with age ($r = .174$, $P < .05$). Neither the composite cloze or completion measures had significant correlations with age, nor did the composite knowledge measure. At the individual scale level, though, significant positive correlations were found for age and knowledge of U.S. history and U.S. literature ($r = .249$ and .234, respectively), and a significantly negative correlation was found for age and biology knowledge ($r = -.164$).

Although not a central question in this study, gender comparisons show that there was a significant advantage to men over women on Gc ($r = -.179$), and for the knowledge composite ($r = -.252$). The latter is mostly attributable to the large gender differences on the technology scale and the U.S. history scale. No significant correlations were found for gender and the composite cloze and completion tests, though Cloze #1 — Biology showed a significant advantage to men ($r = -.171$).

4.3. Cloze and completion tests

Substantial positive correlations were found between all of the ability and knowledge composite measures. Cloze and completion composites correlated with one another at $r = .866$. The cloze composite correlated $r = .766$ with Gc, while the completion composite correlated significantly higher with Gc, $r = .807$ (t-test for the difference of dependent correlations; $t = 1.76$, $p < .05$). A similar pattern of correlations was found with the knowledge composite ($r = .676$ for cloze, $r = .734$ for completion; $t = 2.13$, $p < .05$). That is, the completion composite shared more common variance with both Gc and the knowledge composite in comparison with the cloze composite — suggesting that the additional involvement of memory in the completion test accounted for significantly more variance in both the other composites.

Interestingly, the cloze composite and, to a greater degree, the completion composite appear to act as suppressor variables for the relationship between Gc and Age. That is, while age and Gc showed a small, but significant positive correlation ($r = .174$), partialling out the variance attributed to cloze or completion composites boosted the correlations between Gc and age. Specifically, $r_{\text{Gc,age.cloze}} = .251$, $p < .05$; and $r_{\text{Gc,age.completion}} = .367$, $p < .01$. Of course, given the nature of partial correlations, one could similarly assert that Gc operates as a suppressor variable for the association between age and cloze and completion composites, namely: $r_{\text{cloze,age.Gc}} = -.184$, $p < .05$; and $r_{\text{completion,age.Gc}} = -.332$, $p < .05$. The interpretations of such statistics is that if one partials-out the influence of cloze or completion test performance, the correlation between Gc and age is substantially positive; or, if one partials-out the influence of Gc, the relations between age and the cloze and completion tests are substantially negative. Such results are consistent with the notion that the Gc tests are

more dependent on the recall of previously learned information (Cattell, 1957; Horn, 1968), but that the cloze and especially the completion tests assess more real-time processing of verbal information, which is more likely to be negatively associated with increasing adult ages — especially if, as surmised, working memory is involved in the performance of the cloze and completion tests.

4.4. Gc and cloze/completion

Another way to examine the common variance among the cloze, completion, and Gc composites is to compute multiple regressions, with Gc as the variable to-be-predicted. Together, the cloze and completion tests account for 67% of the variance in Gc ($R^2 = .668$, $p < .01$). When the cloze composite is entered into the equation first, the completion composite accounts for 8.5% additional variance in Gc [$F(1,161) = 42.06$, $p < .01$]. However, when the completion composite is entered into the equation first, the cloze composite only accounts for an additional 2.2% of variance in Gc [$F(1,161) = 10.67$, $p < .01$].

4.5. Knowledge, Gc, and cloze/completion

Similar multiple regression computations can be made for predicting knowledge composite scores, using cloze, completion, and Gc composites. When Gc is entered into the prediction equation first, it accounts for 62% of knowledge variance ($R^2 = .616$, $p < .01$). Adding the completion test composite explains an additional 1.8% of variance in knowledge [$F(1,161) = 8.21$, $p < .01$], while then adding the cloze composite results in no significant increment to prediction [$F(1,160) = .06$]. Reversing the order of cloze and completion entry, placing cloze in the equation after Gc, shows that the cloze composite still adds no significant increment in predicting knowledge over and above the influence of Gc. That is, even though the individual contributions of cloze and completion composites in accounting for individual differences in knowledge are substantial ($r^2 = .457$ and .539, respectively), only the completion composite provides an incremental predictive validity over the Gc composite.

4.6. Structural equation modeling

A confirmatory factor analysis with LISREL allowed the simultaneous estimation of the latent variables. Our model hypothesized four latent variables (see Fig. 1). The model was constructed with traditional measures of Gc as indicators of the Gc factor, four cloze tests as indicators of a cloze factor, four completion tests as indicators of a completion factor and the four knowledge tests as indicators of a knowledge factor. These factors were allowed to correlate. The model fit was good, $\chi^2 = 297.85$, $df = 113$, $p < .001$, Comparative Fit Index (CFI) = .92, Normed Fit Index (NFI) = .87, Non-Normed Fit Index (NNFI) = .90 (see Bentler & Bonett, 1980 for a discussion of these fit indexes). Clearly, the model reflects the fact that all of the measures share substantial common variance. Specifically, the Gc factor is highly associated with the knowledge factor, as are the cloze and completion factors. While the difference is not great, greater commonality is found between the Gc and completion factors than is found between the Gc and cloze factors. The model also demonstrates greater

Table 2
Correlations among Gc, cloze, completion tests, knowledge scales, age, gender, and composite scores

	1	2	3	4	5	6	7	8	9
1. Vocabulary									
2. Information	.665								
3. Word beginnings	.401	.428							
4. Comprehension	.592	.663	.420						
5. Synonyms	.698	.645	.451	.718					
6. Cloze #1	.556	.626	.464	.630	.599				
7. Cloze #2	.584	.626	.446	.670	.666	.741			
8. Cloze #3	.650	.628	.463	.601	.653	.724	.773		
9. Cloze #4	.450	.465	.461	.554	.515	.680	.756	.751	
10. Completion #1	.593	.598	.507	.588	.656	.752	.719	.696	.651
11. Completion #2	.687	.643	.530	.636	.736	.696	.738	.744	.689
12. Completion #3	.674	.658	.502	.642	.703	.751	.734	.776	.670
13. Completion #4	.484	.525	.469	.569	.616	.634	.670	.690	.679
14. U.S. history	.699	.667	.219	.515	.604	.470	.498	.555	.339
15. U.S. literature	.735	.591	.325	.476	.618	.482	.576	.553	.484
16. Biology	.494	.492	.269	.473	.474	.495	.519	.392	.381
17. Technology	.583	.676	.345	.594	.569	.640	.537	.540	.401
18. Age	.360	.173	.044	.114	.069	−.021	.018	.166	.012
19. Gender	−.078	−.204	−.045	−.191	−.189	−.171	−.046	−.087	.074
Composite scores									
20. Gc	.821	.833	.659	.839	.873	.697	.727	.725	.578
21. Cloze Comp.	.625	.654	.511	.690	.684	.878	.912	.906	.889
22. Completion Comp.	.674	.668	.554	.672	.750	.781	.787	.799	.736
23. Knowledge Comp.	.775	.748	.358	.638	.705	.643	.656	.629	.494

Correlations greater than $r = \pm .153$, significant $P < .05$; greater than $r = \pm .205$, significant $P < .01$, two-tailed.

Comp. = composite. Gc = Crystallized intelligence composite.

Cloze #1 and Completion #1: biology. Cloze #2 and Completion #2: U.S. literature. Cloze #3 and Completion #3: U.S. history. Cloze #4 and Completion #4: narrative passage.

Gender coded as: male = 1, female = 2.

10	11	12	13	14	15	16	17	18	19	20	21	22
.809												
.761	.822											
.716	.737	.733										
.451	.514	.622	.441									
.519	.620	.600	.466	.647								
.573	.538	.535	.464	.419	.494							
.586	.568	.626	.509	.642	.454	.580						
−.087	−.031	.052	−.003	.249	.234	−.164	−.060					
−.047	−.010	−.147	−.066	−.242	−.007	−.128	−.437	.010				
.720	.785	.776	.634	.656	.665	.536	.665	.174	−.179			
.786	.799	.818	.749	.519	.584	.501	.591	.049	−.064	.766		
.906	.927	.914	.881	.560	.608	.586	.632	−.019	−.075	.807	.866	
.656	.690	.734	.578	.837	.801	.771	.827	.080	−.252	.785	.676	.734

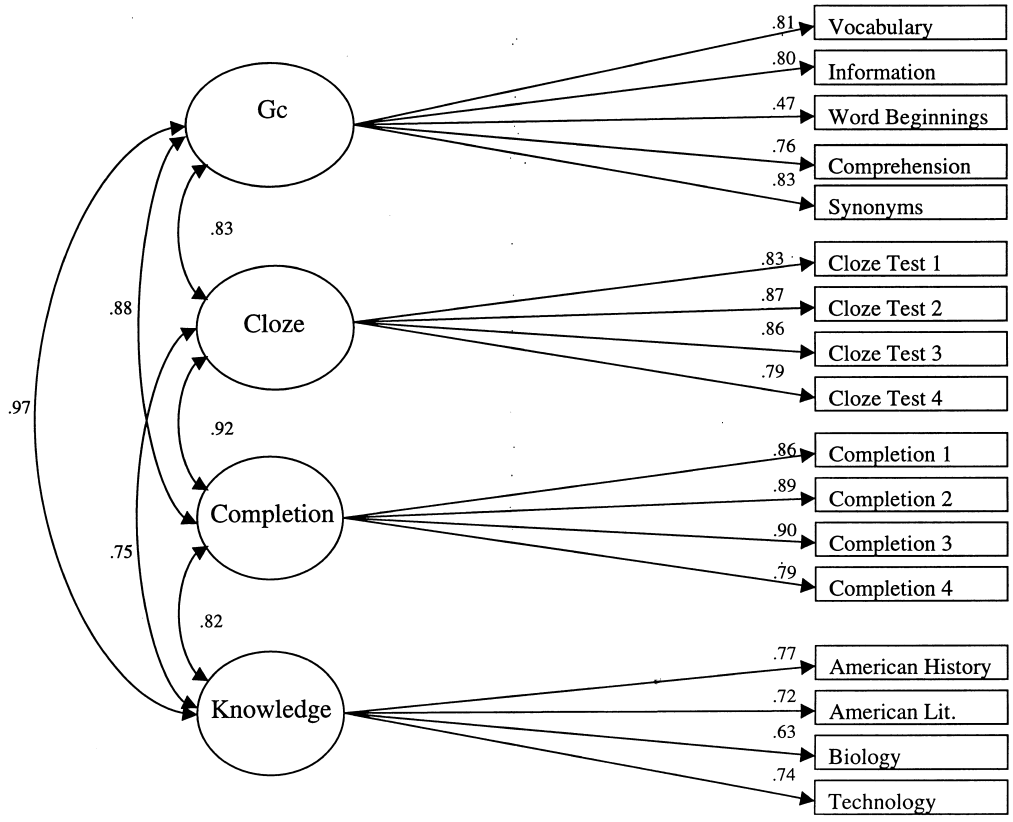

Fig. 1. LISREL confirmatory factor analysis of Gc tests, cloze tests, completion tests, and knowledge scales. Identified factors are Gc, cloze, completion, and knowledge. See text for model fit information.

commonality between the completion and knowledge factors than is found between the cloze and knowledge factors. These findings are supportive of the interpretations offered from the raw correlational analyses. Specifically, the completion tests are more highly associated than the cloze tests are with both the Gc and knowledge factors.

5. Discussion/conclusions

This study has clearly shown that the cloze and completion tests can be useful additions to the assessment of broad crystallized abilities (including verbal, educational, and knowledge components). Even though completion tests should be easier to complete than the cloze tests (given the prior exposure of the full text), the completion tests proved to share more variance than the cloze tests did with both Gc and with the knowledge composites. Moreover, both the cloze and completion tests were found to have suppressor effects in the relationship between Gc and Age, such that removing the influence of individual differences on these tests increased the positive association between Gc and Age.

What underlying processes do these tests actually depend on? Although there was substantial discussion of this issue regarding the completion test from the early 1910s through the 1920s, and some discussion regarding the cloze tests in the 1950s and 1960s, there is much work that needs to be done to put these tests in a current process-oriented framework. In particular, it will be important to determine the role of working memory on performance in both methods of testing, especially given the divergent patterns of age-performance relations found for these tests (essentially zero correlations with age) and Gc tests (positive correlations with age).

While we are not quite as optimistic about the completion test as Spearman (1927) appeared to be, when he called it the single best test of intelligence, we believe that the completion test clearly deserves greater use in the context of assessment of a broad Gc ability, and in the prediction of learning and achievement. We have research currently under way that will assess the relative contributions of both Gf and Gc to completion test performance, and we will also be able to assess whether completion test performance is more highly related to particular knowledge domains (e.g., humanities and civics) over others (e.g., sciences and technology). An optimist would say that the current study shows that, once again, there is great utility to the work of the early modern differential psychologists. Of course, a pessimist would say that in the past 100 or so years, perhaps we have learned less about intelligence than we like to think we have. Regardless, we think that these old tests provide an excellent vehicle for further explorations of the role of abilities in predicting individual differences in learning, especially in the context of providing a means toward assessment of the underlying processes that may give rise to individual differences in Gc and in knowledge.

Given that Gc measures outperform Gf measures in predicting knowledge across the social sciences, humanities, and civics (Ackerman, 2000; Ackerman & Rolfhus, 1999; Rolfhus & Ackerman, 1999), it seems clear that placing the completion test in the foreground (and thus putting abstract reasoning tests as markers for general intelligence in the background), may result in both better predictions of learning and achievement, and a more balanced view of general intelligence than current followers of the Spearman doctrine (e.g., Jensen, 1998).

Acknowledgments

This research was supported by a grant from the National Institute on Aging, (AG16648), Phillip L. Ackerman, principal investigator.

References

Ackerman, P. L. (1996). A theory of adult intellectual development: Process, personality, interests and knowledge. *Intelligence*, *22*, 229–259.

Ackerman, P. L. (1998). Adult intelligence: Sketch of a theory and applications to learning and education. In: M. C. Smith, & T. Pourchot (Eds.), *Adult learning and development: Perspectives from educational psychology* (pp. 143–156). Mahwah, NJ: Lawrence Erlbaum Associates.

Ackerman, P. L. (2000). Domain-specific knowledge as the dark matter of adult intelligence: Gf/gc, personality and interest correlates. *Journal of Gerontology: Psychological Sciences, 55*, 69–84.

Ackerman, P. L., & Rolfhus, E. L. (1999). The locus of adult intelligence: Knowledge, abilities and non-ability traits. *Psychology and Aging*, *14*, 314–330.

Attneave, F. (1959). *Applications of information theory to psychology: A summary of basic concepts, methods and results*. New York: Henry Holt.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, *88*, 588–606.

Campbell, N. A. (1996). *Biology* (4th ed.). Menlo Park, CA: Benjamin/Cummings Publishing.

Carroll, J. B. (1982). The measurement of intelligence. In: R. J. Sternberg (Ed.), *Handbook of human intelligence* (pp. 29–120). Cambridge: Cambridge Univ. Press.

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

Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin*, *40*, 153–193.

Cattell, R. B. (1957). Fluid and crystallized intelligence. In: R. B. Cattell (Ed.), *Personality and motivation* (pp. 871–880). New York: Harcourt, Brace & World.

Cattell, R. B. (1971/1987). *Intelligence: Its structure, growth, and action*. Amsterdam: North Holland (Revised and reprinted from *Abilities: Their structure, growth, and action*, Boston: Houghton-Mifflin).

DiYanni, R. (1994). *Literature: Reading fiction, poetry, drama and the essay* (3rd ed.). New York: McGraw-Hill.

Ebbinghaus, H. (1896–1897). Über eine neue Methode zur Prüfung geistiger Fähigkeiten und ihre Anwendung bei Schulkindem (On a new method for testing mental abilities and its use with school children). *Zeitschrift für Psychologie und Psysiologie der Sinnesorgane*, *13*, 401–459 (translated by Wilhelm, 1999).

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

Ferguson, G. (1956). On transfer and the abilities of man. *Canadian Journal of Psychology*, *10* (3), 121–131.

Hebb, D. O. (1942). The effect of early and late brain injury upon test scores and the nature of normal adult intelligence. *Proceedings of the American Philosophical Society*, *85*, 275–292.

Horn, J. L. (1968). Organization of abilities and the development of intelligence. *Psychological Review*, *75*, 242–259.

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology*, *57* (5), 253–270.

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

Jackson, D. N. (1985). *Multidimensional aptitude battery*. London, Canada: Sigma Assessment Systems.

Jensen, A. R. (1998). *The g factor: The science of mental ability*. Westport, CT: Praeger.

Kelly, T. L. (1928). *Crossroads in the mind of man: A study of differentiable mental abilities*. Stanford, CA: Stanford Univ. Press.

Krueger, F., & Spearman, C. (1907). Die Korrelation zwischen verschiedenen geistigen Leistungsfähigkeiten. *Zeitschrift für Psychologie (Leipzig)*, *44*, 50–114 (cited in Whipple, 1921).

Ohnmacht, F. W., Weaver, W. W., & Kohler, E. T. (1970). Cloze and closure: A factorial study. *Journal of Psychology*, *74*, 205–217.

Rice, A. S., Krout, J. A., & Harris, C. M. (1991). *United States history to 1877* (8th ed.). New York: Harper Collins.

Rolfhus, E. L., & Ackerman, P. L. (1999). Assessing individual differences in knowledge: Knowledge structures and traits. *Journal of Educational Psychology*, *91*, 511–526.

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review*, *103*, 403–428.

Spearman, C. (1927). *The nature of "Intelligence" and the principles of cognition*. New York: MacMillan.

Taylor, W. L. (1953). Cloze procedure: A new tool for measuring readability. *Journalism Quarterly*, *30*, 415–433.

Taylor, W. L. (1954). *Application of cloze and entropy measures to the study of contextual constraint in samples of continuous prose*. Unpublished doctoral dissertation. Urbana, IL: University of Illinois.

Taylor, W. L. (1957). Cloze readability scores as indices of individual differences in comprehension and aptitude. *Journal of Applied Psychology*, *41*, 19–26.

Terman, L. M. (1906). Genius and stupidity. *Pedagogical Seminary*, *13*, 307–373.

Thomson, G. H. (1946). *The factorial analysis of human ability.* Boston: Houghton Mifflin.
Vernon, P. E. (1950). *The structure of human abilities*. New York: Wiley.
Weaver, W. W., & Kingston, A. J. (1963). A factor analysis of the Cloze procedure and other measures of reading and language ability. *Journal of Communication*, *13*, 252–261.
Wechsler, D. (1958). *The measurement and appraisal of adult intelligence*. Baltimore, MD: Williams and Wilkins.
Wechsler, D. (1981). *Manual for the Wechsler adult intelligence scale-revised*. New York: Psychological Corporation.
Whipple, G. M. (1921). *Manual of mental and physical tests: Part II. Complex processes*. Baltimore, MD: Warwick & York.