SELF-REPORTS OF INTELLIGENCE: ARE THEY USEFUL AS PROXY MEASURES OF IQ?

by

DARIA C. LYSY

B.A., University of Western Ontario, 1993
Hons. B.A., University of Windsor, 1994

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ARTS

in

THE FACULTY OF GRADUATE STUDIES

Department of Psychology

We accept this thesis as conforming to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

March 1997

©Daria C. Lysy, 1997
In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Psychology
The University of British Columbia
Vancouver, Canada

Date March 19/97
Abstract

Correlations between single self-ratings of intelligence and IQ are rather small (.20-.25) in college samples. Possible improvements in traditional methods were investigated by employing (a) indirect questions and (b) aggregation. To evaluate these improvements, we compared the validity of aggregated and unaggregated versions of direct measures with four indirect measures: Gough's Intellectual efficiency scale, Hogan's Intellect composite scale, Sternberg's Behavior Check List, and Trapnell's Smart scale. We also compared the performance of a novel self-report measure, Paulhus' Over-Claiming Questionnaire, which shares properties of IQ tests and self-report measures. All measures were administered to two large samples of undergraduates (Ns = 310, 326), who also took an IQ test. Results with traditional self-reports showed that both direct and indirect measures can reliably predict IQ scores but the validity cap appears to be .30 in our competitive college sample. As a rule, the most valid of the traditional items were global characterizations of mental ability; Aggregation benefited indirect more than direct measures. The novel measure, the Over-Claiming Questionnaire, outperformed all other measures with a validity cap of about .50.
# Table of Contents

Abstract ........................................................................................................ ii
List of Tables ................................................................................................ iv
Introduction .................................................................................................. 1
  Improving Traditional Self-Report Measures of IQ ..................................... 2
  Four Traditional Indirect Measures .............................................................. 3
  A New Indirect Self-Report Measure ............................................................ 5
  The Present Study ......................................................................................... 6
Hypotheses ...................................................................................................... 7
Method ............................................................................................................ 8
  Subjects and Procedure ................................................................................. 8
  Instruments ................................................................................................... 8
Results ........................................................................................................... 12
  Descriptive Statistics .................................................................................. 12
  Range of Responses ..................................................................................... 13
  Intercorrelations Among Predictors .............................................................. 14
  Performance of the Traditional Direct Measures ........................................ 14
  Performance of the Traditional Indirect Measures ..................................... 14
  The Two-Factor Categorization of the Traditional Predictors ..................... 15
A Closer Examination of the Traditional Indirect Measures: 
  Empirical Approach ..................................................................................... 16
A Closer Examination of the Traditional Indirect Measures: 
  Theoretical Approach ................................................................................ 17
  Performance of the New Indirect Measure: The OCQ ............................... 18
Discussion .................................................................................................... 19
  Performance of the Traditional Direct Measures ........................................ 19
  Performance of the Traditional Indirect Measures ..................................... 20
  The Content of the Traditional Indirect Predictors .................................... 22
  The Value of Aggregation ........................................................................... 23
  Why Don't Traditional Self-Report Measures Work? ............................... 23
  Performance of the New Indirect Measure: The OCQ ............................... 25
Other Promising Avenues ............................................................................ 26
Conclusions .................................................................................................. 27
Footnotes ...................................................................................................... 28
References .................................................................................................... 30
Appendices ................................................................................................. 44
List of Tables

1. A two-factor categorization of traditional self-report intelligence measures .................. 35
2. Descriptive statistics .................................................. 36
3. Intercorrelations among indirect scales and subscales .......................... 38
4. Correlations of self-report predictors with IQ ........................................ 39
5. Regression of IQ on four traditional indirect predictors ........ 40
6. Correlations of four types of traditional predictors with IQ ............. 41
7. Top 10 item validities from the traditional indirect scales .................. 42
8. Validities of the traditional indirect items:
   Means within content category ............................................. 43
Self-Reports of Intelligence:
Are They Useful as Proxy Measures of IQ?

Can people validly rate their own intelligence? Skeptics argue that such self-reports are hopelessly contaminated with a variety of distortions including self-deception and impression management, and reconstrual. Even people who are free of such distortions—the most forthright and insightful individuals—can find it difficult to evaluate themselves accurately on this elusive trait.

Despite its elusiveness, the concept of intelligence plays a central role in psychological research, particularly in such contexts as educational evaluation, personnel selection and child development. To facilitate such research, considerable effort has been devoted to developing self-report alternatives to cumbersome IQ tests of intelligence. Such approaches have progressed well beyond a simple request to "rate how intelligent you are". Two strategies, in particular, have been recommended. One is the use of indirect assessment to bypass the inevitable defensiveness of a direct request for a self-rating. A second, the aggregation strategy, favors multiple item over single-item measures. This report evaluates these strategies by determining their ability to improve prediction of performance on IQ tests.

Use of IQ tests as the criterion for intelligence ratings has not yielded high validities\(^1\), particularly in college samples. The validities are somewhat higher in observer-ratings than in self-ratings. And outside observers, even with limited familiarity, show reasonable success. Values in the range of .25 to .50 have been found when the judgment is made by spouses (Bailey & Mettetal, 1977), by friends and strangers (Borkenau, 1993), by adolescent acquaintances (Bailey & Hatch, 1979), and by long-term discussion-group colleagues (Paulhus & Morgan, 1997). Except in the case of spouses, however, the achievement of these solid validities required aggregation across multiple observers.

Self-perceptions typically parallel other-perceptions but, to the extent that the trait being evaluated is highly evaluative (e.g., intelligence), the former are noticeably less valid (John & Robins, 1993). Studies using IQ as the criterion have yielded single-item validities of .32 (Borkenau & Liebeler, 1993) and .41
(Reynolds & Gifford, 1996) in broad-band samples. But in college samples, the validities never surpass .30: For example, .26 (DeNisi & Shaw, 1977), .25 (Paulhus & Morgan, 1997), .23 (Sternberg, Conway, Ketron, & Bernstein, 1981, p.49), and .26 (Reilly & Mulhern, 1995).

It is likely that several factors contribute to the modest size of these self-report validities. First, IQ scores will have little influence on self-perceptions if respondents, even those aware of their cognitive abilities, consider them irrelevant to intelligence. Second, self-presentation is likely to operate on self-reports of such a valued quality; Even if self-presentation does not change the ordering of individuals, it will restrict the range of responses to the top few scale options. Third, use of a single rating item virtually guarantees low reliability of measurement. Finally, the validities observed in college samples (.20-.26) suffer an additional handicap--severe restriction in range of the criterion IQ scores. This modest level of validity is the starting point for the present report.

Improving Traditional Self-Report Measures of IQ

The literature contains a number of self-report instruments that show potential as proxy IQ scales, that is, economical substitutes for IQ tests: If valid, such scales have great practical advantages over their traditional counterparts: Rather than running subjects one-by-one in a tightly supervised laboratory setting, researchers can administer such scales quickly to large groups of subjects. Moreover, self-report questionnaires are less threatening than IQ tests and therefore more likely to elicit cooperation. Of course, such advantages are pointless unless validities can be improved over those values cited above. An ideal proxy scale would represent, in effect, a parallel measure showing a validity equal to the reliability of IQ tests, that is, around .90 for the test used in this report (Wonderlic, 1992). Given that standard tests differ somewhat in emphasis, however, a more appropriate upper-limit is the correlation between two well-validated IQ tests, that is, roughly, .80-.85 (Thorndike, 1982). Even that level of association seems unlikely, given that self-report scales use
such a qualitatively-different mode of assessment that tends to be less reliable.

Potential proxy scales in the literature have relied on two strategies, in particular, for improving the validity of self-report measures of intelligence. First is the reduction of evaluation-threat by using subtle, non-obvious questions. The term, indirect measures, is used to describe test formats designed to mask the purpose of the test. The second strategy involves aggregating a set of items to improve reliability.

Four Traditional Indirect Measures

Rather than directly asking "Are you intelligent?", indirect measures pose questions about interests, behaviors, personality, etc. In this report, four such measures were examined: Gough's Intellectual efficiency (Ie) scale, Hogan's Intellect composite scale, Sternberg's Behavior Check List (BCL), and Trapnell's Smart scale. All four have shown some validity in predicting criterion measures of intelligence. Although all have an indirect format, the rationale for each is rather different.

Gough's Intellectual efficiency (Ie) scale. In the first such effort, Gough (1953) developed a set of self-report items for use as a proxy measure of intelligence. He administered a pool of items assumed to tap aspects of personality associated with intelligence. Those 52 items correlating most highly with an IQ test in a sample of high school students were assembled and labeled the Intellectual efficiency (Ie) scale. In four cross-validation studies, Gough reported a mean validity of .47. In the four educated samples reported in the latest manual, however, the median validity is .29 (Gough, 1996).

As for most tests derived from contrasted groups, the Ie items are rather heterogeneous: Topics included self-confidence, neuroticism, intellectual abilities and interests, and social skills. Thus the vast majority were subtle indicators, that is, they lacked face-validity as indicators of intelligence. As such they are less likely to trigger self-presentation. Appendix A displays the Intellectual efficiency scale.

Hogan's Intellect composite scale. Welsh (1975) coined the term intellectance to denote the "cognitive and interpersonal style that causes people to be perceived as bright". Hogan and
Hogan (1992, p. 12) followed this peer-perception notion of Intellect in assembling a set of items. A factor analysis revealed two factors: One was labeled Intellectance: "the degree to which a person is perceived as bright, creative, and interested in intellectual matters". The other factor was labeled School Success: "the degree to which a person seems to enjoy academic activities and to value educational achievement for its own sake".

Intellectance items refer to science ability, curiosity (about the world), thrill seeking, interest in intellectual games and generating ideas (ideational fluency), and interest in culture items, while School Success items concern education (being a good student), math ability, good memory, and enjoyment of reading. Observers tend to see high scorers on the Intellectance scale as "imaginative, inventive, and quick-witted, but easily bored and inattentive to detail" whereas low scorers tend to be "unimaginative, narrow, tolerant of boredom, and not needing much stimulation". In contrast, high scorers on the School Success scale are seen as "foresighted, thorough, and painstaking" whereas low scorers are seen as "touchy, restless, and impulsive" (Hogan & Hogan, 1992, p.40). Appendix B displays the Intellect composite scale.

Sternberg's Behavior Check List (BCL). As part of his investigation into conceptions of intelligence, Sternberg (1988, p.238) developed the Behavior Check List (BCL)--a list of 41 behaviors that lay judges associated with intelligence. Factor analyses indicated three clusters of items labeled Problem Solving (PS), Verbal Ability (VA), and Social Competence (SC). Correlations with an IQ test jumped from .24, when employing self-perceived intelligence, to .52 with the full-scale BCL. These results were later replicated by Cornelius, Kenny, and Caspi (1989).

Sternberg recommended the BCL as a valuable supplementary measure of intelligence for a number of reasons. Compared to providing a global assessment of their ability, subjects should feel less-threatened by rating specific behaviors and, accordingly, be more accurate. Aggregation of a large set of these specific behaviors could then yield a maximally valid self-
report. Finally, the BCL could cover aspects of intelligence not measured by IQ. Appendix C displays the Behavior Check List.

**Trapnell's Smart scale.** The four-item Smart scale measures self-appraised intelligence via simple trait descriptive statements of high face validity (Trapnell, 1994). The content of three of the Smart scale items was based on the assumption that range restriction in self-ratings due to desirable responding can be reduced through the use of extreme qualifiers (e.g., very, extremely, exceptionally) and by shifting the implied locus of evaluation from the self to others (e.g., "I'm considered to be..." in place of "I am..."). A fourth item assessed self-reported school grades, based on the assumption that grades provide an indirect but objective index of mental ability that can be recalled and self-reported fairly accurately. The Smart scale correlated .33 with an IQ test in a college sample (Trapnell & Scratchley, 1996). Appendix D displays the Smart scale.

**A New Indirect Self-Report Measure**

**Paulhus' Over-Claiming Questionnaire (OCQ).** Designed to measure both intellectual ability and intellectual enhancement, the Over-Claiming Questionnaire is a comprehensive self-rated knowledge task that functions as an objective test because accurate knowledge can be measured. Thus it shares properties of an IQ test and a self-report measure. Academic and everyday knowledge are systematically sampled with a wide-ranging set of 90 general knowledge items to which respondents rate their familiarity on a scale ranging from 0 (never heard of it) to 6 (know it very well).

Categories include historical names and events, books and poems, authors and characters, social science and law, physical sciences, and popular culture names (see Appendix E). For every four existent items, one nonexistent item is included, thus within each category, three out of every 15 items are foils, that is, they do not actually exist. Hence any degree of claimed knowledge about them constitutes overclaiming. The three foils for each category were selected to closely resemble the 12 existent items and thus appeared plausible to a non-expert. In
total, overclaiming is possible on 18 items spread across a variety of topics.

Signal detection analysis is applied to the familiarity ratings to index **Discrimination** (the ability to distinguish existent from nonexistent items) and **Bias** (tendency to overclaim familiarity). Relevant to the present study is **Discrimination**, which is indexed by the number of hits relative to the number of false-alarms. An accurate individual, then, is not the one achieving the most hits, but the one showing the most discrimination in choosing between existent and nonexistent items.

Paulhus and his colleagues have examined the properties of the **OCQ**. Results are reported in detail in Paulhus (1992). In the scale development studies, the reliabilities of the **Discrimination** index range from .75 to .94 and those of the response **Bias** index range from .50 to .90, and the **Discrimination** index correlates .44 with IQ.

**The Present Study**

Are self-reports of intelligence useful proxy measures of IQ? The present study will examine this question by comparing their validities. The traditional self-report measures differ with respect to directness (direct vs. indirect) and aggregation (single-item vs. composites). Thus each falls into one of the four categories of a 2 x 2 table (see Table 1). The first category--single-item direct measures--is represented by the adjectives "intelligent" (Sample 1) and "clear-thinking, intelligent" (Sample 2). The composite direct measure combined a set of four conceptually similar items. The four traditional indirect measures were Gough's **Ie**, Hogan's **Intelllect**, Sternberg's **Behavior Check List (BCL)**, and Trapnell's **Smart** scales. To evaluate the fourth category of measures--single-item indirect--the average item validity for each indirect measure was calculated.

------------------

Insert Table 1 about here

------------------

Within the composite indirect category, there is a specific interest in the comparative validity of the four traditional
measures. Although there is some evidence for the validity of each scale, they have never been pitted against one another in predicting a common criterion. A comparative validity study by non-partisan researchers should provide much more convincing evidence than that offered by the authors of individual scales.

Nor have these measures been compared with the new indirect self-report measure, the Over-Claiming Questionnaire. The latter differs from the other self-reports in more closely resembling an objective test. Thus comparing the validity of the Discrimination index of the OCQ will prove quite interesting and should provide some insight into the nature and possible similarity between self-reports of intelligence and of knowledge.

All measures were administered to two large and diverse samples of undergraduates. The criterion for validity of the self-reports was the Wonderlic Personnel Test, a speeded IQ test that compares favorably with longer IQ tests. The present analyses focused on comparing the validities of the four categories of traditional measures and the Discrimination index of the OCQ via correlation and regression techniques. To put all these validities in perspective, the performance of ideal proxy scales was also estimated.

Hypotheses
The first hypothesis, that aggregation will enhance validities of traditional scales, is based on the basic psychometric principle of aggregation (e.g., Thorndike, 1982). Hypotheses about the value of indirectness are more difficult. Competing theories from the literature suggest opposite outcomes. Many researchers argue that the simple, global, clear (i.e., direct) questions tend to be more valid (e.g., Burisch, 1984). On the other hand, the authors of the indirect measures would argue for the benefits of their indirect scales—and there is evidence for both. Accordingly, a clear hypothesis about indirectness cannot be voiced.

Hypothesizing which indirect measure will perform best is also difficult. The highest validity reported in a college sample is the .44 found by Paulhus (1992) using the OCQ Discrimination index. Therefore, the second hypothesis is that the OCQ will show the highest validity in the present samples.
Method

Subjects and Procedure

Data were collected from a total of 636 undergraduate students at a large northwestern university. Sample 1 comprised 310 students (95 males; 208 females; 7 did not specify) enrolled in an introductory psychology course. Sample 2 comprised 326 students (87 males; 205 females; 34 did not specify) enrolled in a second year social-personality psychology course. Approximately 55 percent of the both samples were liberal arts majors, 20 percent science or engineering majors, and 15 percent business school majors. All participated for extra marks.

For both samples, subjects were first asked to complete a self-report inventory in group sessions: It included all the direct self-ratings of intelligence. Later, a set of indirect measures of intelligence and the new indirect measure were distributed in a take-home package, which subjects were asked to complete privately and return for experimental credits. Finally, the IQ test was administered in a separate, supervised session.

Instruments

Direct measures. A number of intelligence-related items were included in the context of a larger personality inventory. They were selected a priori for their conceptual relevance to intelligence. In Sample 1 they included the following four items: "Is intelligent", "Is ingenious, a deep thinker", "Is smart", and "Is not exceptionally gifted at academic things" (Reverse coded); In Sample 2 the direct items included: "Is clear-thinking, intelligent", "Wants things to be simple and clear-cut", "Is clever, sharp-witted", and "Enjoys thinking about complicated problems". Subjects were asked to rate their agreement with these items on a scale ranging from "1" ("Disagree strongly") to "5" ("Agree strongly"). For special consideration the most face-valid item ("Is intelligent" in Sample 1; "Is clear-thinking, intelligent" in Sample 2) was identified.

To evaluate the utility of aggregating items, the four items judged by three raters to be the most face-valid indicators of intelligence were combined. Although the scale items differed somewhat in the two samples, the similarity of correlates (see
Results section) suggests that the two direct composites measured a similar construct.

Indirect Measures. Given the review of the four indirect measures in the introduction, only a flavor of each one will be provided here. The Intellectual efficiency subscale of the California Psychological Inventory (Ie) (Gough, 1953) included 52 True-False statements. The content includes personality--related items, ranging in content from beliefs (e.g., "Success is a matter of will power") and interests (e.g., "I like to read about history") to bizarre items about experiences (e.g., "I have never seen a vision").

Portions of two subscales of the Hogan Personality Inventory (Hogan & Hogan, 1992) were also included to represent the Intellect factor. All items are in True-False format and were developed from a peer-perception view of intellect. For reasons of convenience and space, the selection was limited to 15 items from the Intellectance subscale and 7 items from the School Success subscale. Examples are "I’m good at inventing games, stories, and rhymes" from the Intellectance subscale and "As a child I was always reading" from the School Success subscale.

The Sternberg Behavior Check List (BCL) consists of short, specific, behavioral descriptions originally selected by lay judges as prototypical of intelligent people (Sternberg et al., 1981). The final 41-item version provided by Sternberg (1988, pp.238-239) was used. Subjects rated from 1 (low) to 9 (high) the extent to which each item was an 'accurate self-description'. The BCL includes three subscales: the 13-item Verbal Ability subscale (e.g., "Speaks clearly and articulately"), the 15-item Problem Solving subscale (e.g., "Makes good decisions"), and the 13-item Social Competence subscale (e.g., "Responds thoughtfully to others' ideas").

The 4-item Smart scale measures self-appraised intelligence via simple trait descriptive statements of high face validity (Trapnell, 1994). The items are: (1) I'm considered exceptionally or unusually intelligent. (2) I'm considered a very "brainy", scholarly person.(3) I'm considered extremely "gifted" or talented at academic things and (4) My school grades have usually been near the top of every class.
The Over-Claiming Questionnaire. The OCQ exploits a sophisticated methodology capable of being a proxy IQ test. Respondents are asked to rate their familiarity with a wide range of people, books, events, etc. The 90 items were culled from comprehensive lists provided by Hirsch (1988) in the appendix of his book, Cultural Literacy and partitioned into six categories (see Appendix). Because 20 percent of the items are fictitious, signal detection statistics (Swets, 1964) can be used to separate accuracy from bias. In this approach, every response falls into one of four categories: (1) hits: claiming existent items are familiar, (2) false-alarms: claiming nonexistent items are familiar, (3) misses: claiming existent items are unfamiliar and (4) correct rejections: claiming nonexistent items are unfamiliar.

Signal detection analysis exploits all of the data in the calculation of separate indexes for Discrimination and Bias. Discrimination is indexed by the number of hits relative to the number of false-alarms, and thus a high-scorer has shown ability in discriminating between existent and nonexistent items. To calculate the formula, the mean number of hits is divided by the mean number of false alarms plus one, respectively for each category. Next, the Discrimination index is calculated by finding the mean for all six categories. Bias is calculated by adding together the mean number of hits and the mean number of false alarms then dividing by two, separately for each category, and then finding the mean of the six categories to determine the overall claim rate. A respondent with a high Bias score has an indiscriminant tendency to say "Yes", i.e., claim a lot of knowledge. The present study is solely concerned with the Discrimination index.

Examples of existent items and foils, i.e., nonexistent items, are "The Devine Comedy" (existent) and "Windermere Wild" (nonexistent) from the "books and poems" category and "Clara Barton" (existent) and "El Puente" (nonexistent) from the "authors and characters" category. In a college sample, the Discrimination parameter correlated .44 with scores on an IQ test (Paulhus, 1992). Considering that it was a college sample, the validity is quite promising.
Objective Measure (IQ test). The 12-minute Wonderlic Personnel Test was chosen to assess IQ. It is a short-form test of general cognitive ability, that is, "the level at which an individual learns, understands instructions and solves problems" (Wonderlic, 1992, p. 5). Included are items sampled from verbal, quantitative, and analytic domains. Although a time-limit is imposed, the Wonderlic behaves more like a power test than a speeded test because the items are presented in ascending order (McKelvie, 1994).

The Wonderlic is the most popular IQ test in applied settings because of its ease of administration and comprehensive norms combined with ample reliability and validity evidence. Expert reviews have been highly favorable (see Aiken, 1996; Hunter, 1989; Schmidt, 1985; Schoenfeldt, 1985).

The Wonderlic shows test-retest reliabilities ranging from .82 to .94 (Dodrill, 1983; Wonderlic, 1992), and alternate-form reliabilities ranging from .73 to .95 (Wonderlic, 1992). These findings are based on adult working populations, however. Because of restriction of range of ability, college samples should yield lower standard deviations and therefore, lower reliabilities. McKelvie (1989) reported a high internal consistency of .87 (odd-even split-half correlation) in a college sample. The fact that reliability is not increased by relaxing the time requirement (McKelvie, 1994) indicates that the time limit does not inflate the estimate.

In support of concurrent validity, the Wonderlic shows correlations above .80 with longer IQ tests such as the WAIS-R (Dodrill, 1981; Wonderlic, 1992). In fact, Dodrill (1981, p. 668) reported that the Wonderlic IQ scores were within 10 points of the WAIS Full Scale IQ scores in 90% of the cases. Of particular note for this report is the fact that correlations are high with measures of both verbal and quantitative abilities (Wonderlic, 1992). Previous studies in college populations have also shown useful predictive validity for college grades (McKelvie, 1994), performance tests (Kennedy, Baltzley, Turnage, & Jones, 1989), and supervisory rankings (Wonderlic, 1992).
Results

Descriptive statistics

Means, standard deviations, ranges, and reliability coefficient alphas are presented in Table 2. The values of these statistics in the two samples are virtually identical. Alpha values for the full scales and subscales are generally quite acceptable, ranging from .61 to .93 in Sample 1 and .55 to .92 in Sample 2. The reliability of the single-item “intelligent” was estimated from the mean intercorrelation of the four global items from the direct composite.

Although the OCQ has a 7-point rating scale, signal detection analysis requires dichotomous coding for hits and false alarms. Both are accumulated when a subject claims familiarity with any response greater than zero. Because each category of the OCQ contains 12 existent items and 3 foils, i.e., nonexistent items, a subject’s mean hit proportion is the mean number of hits in each category divided by the number of possible hits, that is, 12; The mean false alarm proportion is the mean number of false alarms in each category divided by the number of possible false alarms, that is, 3. These statistics are presented in the “item mean” column of Table 2. Mean item familiarity ratings are presented in Appendix F.

Note the means and standard deviations (SDs) for the Wonderlic IQ test, which are not presented in the table. Sample 1 and 2 means (25.5, 26.3) were only slightly higher than the manual norms for college students (Wonderlic, 1992, p.38). SDs (4.41, 4.72), however, were substantially lower than the manual norms of 5.73 for college students. For comparison, note that the norms computed on a representative adult working population (Wonderlic, 1992, p.38) exhibited a substantially lower overall mean (21.6) and higher SD (7.1).

The alpha reliabilities for the Wonderlic were calculated in two ways. Although not ideal, internal consistency was estimated directly in our sample with the odd-even split half-reliability used by McKelvie (1989). These values were .65 and .71. A
second calculation involved extrapolating from the appropriate reliability estimates (.90) taken on the broad norm sample (Wonderlic, 1992). Applying the formula from Gulliksen (1967, p.124), to the reduction in standard deviation from 7.12 to 4.41 and 4.72, the alphas in the present samples were estimated to be .74 and .77. Using either estimation formula, the reliabilities in the present college samples were noticeably lower than in the population, but certainly within the useful range for research instruments. It can be expected that the present validities, in turn, will be correspondingly lower.

Range of Responses

In the introduction, it was noted that the strong tendency for respondents to claim high levels of intelligence tends to restrict the range of responses and limit potential correlations (McCrae, 1990; Thorndike, 1982). Note that the SDs shown in Table 2 were calculated on the subject means, rather than calculating the means of the item SDs. Given that the latter figures are more relevant to whether or not subjects were using the entire range of the rating-scales, these were the figures calculated.

Recall that the direct items were measured on 5-point rating scales: The exact distribution of responses was (0, .05, .25, .40, .30) across the two samples. The SD for the single direct item was only .77 and .82 in Samples 1 and 2, respectively. For the four items of the composite direct scale, the average SDs were still small: .67 (Sample 1) and .69 (Sample 2). In short, the direct items did show some restriction in range.

For the indirect measures, the means of the item standard deviations and ranges were not relevant for the two True-False scales (Gough's Intellectual efficiency and Hogan's Intellect composite) and were therefore not calculated. The same argument applies to the OCQ because the calculation of hits and false alarms are now the subject of concern rather than the original responses. The other two traditional indirect measures were administered in identical 9-point response format, but the variation of the Smart scale items was noticeably greater than the Behavior Check List (BCL) items. The average standard deviations were 1.54 and 1.65 for Trapnell's Smart scale, and .82
and .75 for the BCL. The average range of the items of the Smart scale was fully 8.00 in both samples, higher than that for the BCL, 4.63 and 4.10.

**Intercorrelations Among Predictors**

The matrix of intercorrelations among the traditional indirect scales, subscales, and the new indirect measure, is presented in Table 3. Note that the four traditional indirect measures (not including subscales) and the new indirect measure intercorrelate positively but only modestly, with correlation coefficients ranging from .08 to .56 (Sample 1) and from .10 to .47 (Sample 2). Sternberg's subscales intercorrelated quite strongly, with correlation coefficients ranging from .67 to .77 (Sample 1) and from .65 to .68 (Sample 2), while Hogan's subscales intercorrelate modestly, with correlation coefficients of .25 (Sample 1) and .35 (Sample 2).

Insert Table 3 about here

**Performance of the Traditional Direct Measures**

Table 4 contains the validities, that is, the correlations of all self-rated intelligence measures with IQ. The baseline validity is that of the single self-rated intelligence item: These values were .20 (Sample 1) and .23 (Sample 2). The corresponding validities for the composite direct measure were slightly higher: .24 (Sample 1) and .26 (Sample 2).

Insert Table 4 about here

This small improvement was disappointing. The Spearman-Brown formula\(^{10}\), for instance, would predict a value of .33 for a 4-item scale of .23. Apparently, the validities of the additional items did not parallel those of the original item (intelligent). Despite the best effort to select conceptually similar items, aggregation provided only modest improvements in the validity of direct measures.

**Performance of the Traditional Indirect Measures**

The ability of the four traditional indirect measures to predict IQ scores was examined in two ways. First the validities
of each predictor were calculated and compared; then a regression analysis was performed to determine which of the predictors made independent contributions.

Correlations. From Table 4, it can be seen that all four traditional indirect scales achieved significant validities in both samples. Although two IQ outliers were removed from Sample 1 (one high; one low), the pattern of validities across the two samples was remarkably consistent.

Of the four traditional indirect measures, Gough's Intellectual efficiency scale performed best, with validities of .20 (Sample 1) and .34 (Sample 2), followed by Trapnell's Smart scale, with validities of .24 (Sample 1) and .25 (Sample 2). Although not as successful overall, the other measures each offered a successful subscale: Hogan's School Success performed well at .19 (Sample 1) and .27 (Sample 2), and so did Sternberg's Verbal Ability subscale, at .24 (Sample 1) and .18 (Sample 2).

Regression Analyses. To determine whether the traditional indirect measures made independent contributions in predicting IQ, a regression analysis was conducted in each sample (see Table 5). When IQ score was regressed on all four traditional indirect scales (forced entry) in both samples, Gough's Intellectual efficiency scale and the Smart scale showed significant weights across both samples. Hogan's Intellect composite was also significant in one sample. The resulting variance accounted for by the four traditional indirect measures was 10% (Sample 1) and 16% (Sample 2).

A follow-up set of regression analyses was conducted to determine the combined predictive power of each set of subscales. When the three Sternberg subscales were force-entered, they accounted for a total of 6 and 4 percent of the variance in the two samples. A similar forced-entry with Hogan's subscales accounted for a total of 7 percent of the variance in both samples.
The Two-Factor Categorization of the Traditional Predictors

Table 6 summarizes the key data for this report by displaying the mean validities of the four categories of traditional measures of self-report intelligence. The performance of direct measures can easily be compared with those of indirect measures for both single items and the aggregated scales.

---

Insert Table 6 about here

---

The entries in Table 6 are as follows: The single-item direct validities are the correlations of IQ with the single item, "intelligent" (Study 1) or "clear-thinking, intelligent" (Study 2); The single-item indirect validities are mean item validities across all 119 items of the four traditional indirect scales. The aggregated direct entries are the validities of the 4-item direct scales. Finally, the aggregated indirect validities are the mean scale validities across all four traditional indirect measures.

In the case of the traditional direct measures, aggregation boosts the validities from .20-.23 to .24-.26. Such minimal increases (.03-.04) contrast dramatically with increases around .30 predicted by the Spearman-Brown formula.

For the indirect scales, however, Table 6 reveals a dramatic effect for aggregation. Here the comparison is between the mean of the 119 item validities (.07) and the mean of the four full scale validities: Broken down scale by scale, aggregation raised the validities from .05 to .20, .20 to .24, .05 to .08, and from .11 to .20 for Intellectual efficiency, Smart, Intellect, and the BCL, respectively, in Sample 1 and from .07 to .34, .21 to .25, .09 to .13, and from .06 to .13 in Sample 2. In short, all of the traditional indirect scales benefited from aggregation.

Note that the Spearman-Brown predictions for the aggregated validities are .19, .32, .16, and .30 for the four traditional measures, respectively, in Sample 1 and .24, .33, .25, and .21 in Sample 2. Thus it appears that only the Intellectual efficiency scale benefited from aggregation as much as, or more than, the Spearman-Brown formula would predict. The other three scales
fell short of the potential achievable if each item had been as valid as the mean item.

A Closer Examination of the Traditional Indirect Measures: Empirical Approach

Top Performing Items. Table 7 presents the 10 best performing items from the traditional indirect scales, as defined by consistently good validities. Every measure is represented in the top 10. The fact that the Intellectual efficiency (Ie) scale has the largest number of representatives probably derives from the fact that it contains the most items. It is noteworthy that the items with the highest validities are the intelligence- or ability-related ones. After all, the rationale behind the creation of these indirect measures was that indirect items should exceed the validity of the more blunt, direct items such as "intelligent" or "smart".

All four Smart items performed well in both samples, and two made the top 10. Note that the top Intellect items are from the School Success subscale and both concern reading. Also note that the top items of the Behavior Check List both come from the Verbal Ability subscale.

Summarizing the contents of Table 7, then, it appears that the top-performing items in the traditional indirect scales were either (a) direct ability-related items, (b) indirect items about ability (i.e., Smart scale), and (c) items about reading behavior. If these ten items are combined into a new "best items" composite scale, the correlation with IQ is .34 in Sample 1 and .38 in Sample 2.11

A Closer Examination of the Traditional Indirect Measures: Theoretical Approach

Having discovered that the best performing items of the Gough, Hogan, and Sternberg scales were, in fact, the more direct and ability-related items, the items of these indirect measures were categorized theoretically. Four a priori categories were considered: mental abilities, personality-related, behaviors,
and interests. Two judges showed 95% agreement on classification.

For each category, the mean validities were calculated and presented in Table 8. Mean item validities were .14 (Sample 1) and .12 (Sample 2) for the ability items. Means for next three categories (personality, interests, and behaviors) were in the .03 to .06 range—all substantially lower than the ability-related items.

Best of all, however, was the set of items addressing reading habits: "I read at least ten books a year" from the Ie and Intellect scales, "As a child I was always reading" from Hogan’s Intellect composite scale, and "Reads widely" and "Sets aside time for reading" from the BCL. In fact, these reading items showed exceptional mean validities of .19 (Sample 1) and .18 (Sample 2).

In sum, it appears that items directly related to mental ability and items about reading habits outperformed the other item content categories in predicting IQ. The other categories of items show positive, but low, item validities. Nonetheless, they can be aggregated, as in the case of the Ie scale, to reach a reasonable level of validity.

Performance of the New Indirect Measure: The OCQ

The validity of the Discrimination index of the Over-Claiming Questionnaire is displayed in Table 4. The OCQ clearly outperforms all other measures in both samples. This result might be expected, given that the OCQ shares some resemblance with objective tests. Subjects who show the most accuracy in discriminating between existent items and foils scored higher on the IQ test.

At the same time, the OCQ is more like a self-report. Therefore, its correlations with other self-reports are of interest. Its modest correlations with the traditional indirect measures were already reported in Table 3. The correlation between the OCQ and the direct, self-report intelligence composite was also modest: .23 in Sample 1 and .12 in Sample 2.
Out of interest's sake, an additional regression was performed including OCQ with the four traditional indirect measures. Simply put, when the five predictors were entered, OCQ Discrimination came out as the best predictor in both samples. In fact, it occluded the other measures.

Discussion

The present study set out to evaluate whether IQ can be measured by proxy. That is, can the handy self-report format be used as a substitute for a cumbersome IQ test? Because the validity of a single self-rating of intelligence has not proved adequate, researchers have advocated a number of strategies for improving validity, namely, indirectness and aggregation. Aggregation did indeed enhance the validities of the traditional scales, as hypothesized. The present results indicate that aggregated, direct measures were the most effective, but none of the traditional measures could consistently exceed .30. The second hypothesis was also supported: The OCQ indeed showed the highest validity in the present samples.

Performance of the Traditional Direct Measures

The present study began by establishing the validity of the baseline, that is, a single face-valid self-rating of intelligence. In two large samples, the single-item showed validities of .20 and .23—values that are typical of previous studies. Some studies have reported higher validities but most of those were high-school or other samples with a wide range of talent. Competitive college samples such as the present ones, suffer from a restricted range of ability, which limits potential validity values. In any case, such modest baseline values left plenty of room for improvement via aggregation.

The empirical benefits of aggregation were evaluated by pooling the single item with other intelligence-related items in a composite direct measure. An improvement in validity was observed with the addition of the 3-4 items most synonymous/antonymous to "intelligent", namely, smart, clever, simple, and not gifted. The lack of improvement beyond four items suggests that further aggregation added more noise than valid variance. The battery of items contained every conceivable item related to
mental ability. But few of these were able to capture that facet of self-perception linked to IQ.

But even the observed validity of the 4-item composite (.26) did not match the validity predicted from the Spearman-Brown formula (.33) based on the validity of the original item "intelligent". Clearly, that item plays a special role, both conceptually and empirically.

Performance of the Traditional Indirect Measures

Indirect measures promised to surpass the performance of direct measures by providing a less threatening or evaluative, assessment atmosphere. In terms of predicting IQ, the promise was not fulfilled in this data. Given that each indirect measure raised different issues, however, they will be considered one-by-one.

Gough Intellectual efficiency (Ie) scale. Recall that the Ie scale was constructed in a contrasted-groups fashion by selecting California Psychological Inventory items that correlated with an IQ test (Gough, 1953). Given that it was developed decades ago on California high-school students, its success in a contemporary Canadian college sample--validities of .20 and .34--might be considered remarkable. Nonetheless a full 52 items were required to achieve those full-scale validities because the mean item validity was low. Of course, True-False items are expected to show lower validities (but faster administration times) than corresponding Likert items.

Contrary to the original intent of the Ie, however, it was primarily the mental-ability items that correlated with IQ, and these items tend to be quite direct in nature. Originally, a number of confidence and adjustment items were selected because they correlated with IQ. Those associations were not sustained in the present results. The top items, seen in Table 7, focused directly on ability. Why the confidence and adjustment item validities did not replicate is difficult to say. Perhaps IQ test performance is more influenced by adjustment in high-school than in college students.

Perhaps the four-decade gap in culture is somehow responsible. Even across a 10-year time-span, Paulhus and Landolt (1994) found that rather different people were nominated when subjects were
asked "who is most intelligent". It is suspected that when criterion-groups rather than rational methods are used to develop scales, items measuring temporary societal influences are more likely to intrude.

**Sternberg's Behavior Check List.** As a unit, the BCL showed only modest predictive efficacy. One of the subscales—the Verbal Ability subscale—was effective. Sternberg et al. (1981) found the same pattern. Detailed item analyses in the present study revealed that the high-validity items carrying the subscale were those concerning mental-ability and items about reading habits.

Although the ability of the BCL to predict IQ was not impressive for a 41 Likert-item measure, attention must be called to its original purpose. Sternberg intended the BCL, not as a proxy for IQ tests, but as a supplementary measure: It was to be administered along with an IQ test to tap components of intelligence that IQ tests were not capable of measuring (Sternberg, 1988, p.239). From this perspective, a high correlation with IQ would not be expected.

This perspective on the BCL is consistent with Sternberg's longstanding position that IQ tests measure only a limited part of lay conceptions of intelligence. Recently, this notion has been followed up in work on "non-test intelligence" (Lysy & Paulhus, 1996). By partialing IQ and self-presentation out of self- and peer-ratings of intelligence, a self-residual and a peer-residual was formed indexing that part of intelligence that is "beyond IQ". The residuals were then correlated with a battery of personality and interest measures. The top correlates of the self-residual were self-rated conscientiousness and openness, self-esteem, Intellectual efficiency scale, and the Smart scale, while the top correlates of the peer residual were peer-rated conscientiousness, openness, physical attractiveness and athletic ability. The different correlates of self and peer suggest that "non-test intelligence" is largely a perceiver-dependent idiosyncrasy. There was, however, a small overlapping component indicating that self and others systematically misattribute intelligence to those who are conscientious and
open. This component may be that facet of "true intelligence" that is not represented in IQ tests.

Hogan's *Intellect* composite scale. As a whole, the *Intellect* composite showed only a modest ability to predict IQ scores. Obscured in this overall figure, however, is the fact that the two subscales showed dramatically different validities. Recall that *Intellectance* subscale was designed to capture unconventional, creative conception of intellect whereas *School Success* was aimed at the more conventional goal-oriented conception of intellect. The present results support this distinction in that *School Success* was a distinctly better predictor of IQ. Moreover, the efficacy of the latter was underestimated by employing only a 7-item version.

Using the Spearman-Brown formula, the predicted validity of a 21-item version would be .45--better than any of the mean validities obtained in the present data. Of course, the Spearman-Brown prediction would hold only in the unlikely event that 14 equally-valid items could be added.

*Trapnell's Smart* scale. The newest traditional scale in the study, Trapnell's (1994) *Smart* scale, performed well. It was designed to reduce range-restriction in two ways: (1) by diminishing the desirability of claiming the item and (2) by shifting the implied locus of evaluation from self to others. As intended, the *Smart* scale did show a reduced range restriction: Subjects utilized almost the entire range of the 9-point scale--noticeably more than the range of Sternberg's *Behavior Check List*.

The *Smart* scale is certainly efficient, requiring only four-items to match or even out-perform the other traditional indirect measures. It is now evident, however, that the success of the *Smart* scale did not derive from its indirect nature. Direct composites with four items of similar content (*smart, clever, etc.*) worked just as well as the *Smart* scale. Therefore its success was more likely a function of content rather than a strategic contextualizing of the items.

**The Content of the Traditional Indirect Predictors**

This scale-by-scale analysis of successful items has clarified the source of their success. Although the four traditional
inventories derived from four dramatically different domains, the successful items within each were almost entirely ability-related. Of course, the direct measures were designed to address ability directly. But in the case of indirect measures, it is certainly ironic that the direct items work best, given the rationale behind the first of this study's strategies. Fortunately, this finding refutes the possibility that the lower validities overall for indirect measures resulted from their inclusion in a later test battery administered later than were the direct measures.

Out of all remaining item-content areas, the only one yielding consistently high validities was an interest in reading. Why a lifelong enjoyment of reading is associated with achieving high scores on IQ tests is not clear. Many educational psychologists argue that reading behavior permanently boosts mental abilities (Rayner & Pollatsek, 1989) and is therefore encouraged. Of course, other causal sequences are possible. High intelligence might make reading more enjoyable. Or third variables such as social class or openness to experience might nurture both.

The Value of Aggregation

Administration of the single item "intelligent" is certainly efficient given the practical costs of adding more items to a test battery. And, across all items administered, it was the most consistently valid. Nonetheless, for both traditional direct and indirect measures, aggregation improved validities only up to five items. Beyond that, returns were marginal. Apparently, the items linked to IQ have a limited semantic scope. The fact that the Intellectual efficiency scale benefited most from aggregation suggests that this strategy aids True-False more than Likert-item scales. It is understandable that dichotomous items, though potentially as valid as Likert items, require more aggregation because of lower item reliability.

Why Don't Traditional Self-Report Measures Work?

Can self-reports be used as proxy measures of IQ in college samples? The present data suggest that traditional measures are not satisfactory. Given that the validity of an ideal proxy measure would be upwards of .55 in college samples, the
validity cap of .30. The best available measures, as well as the most highly touted improvement strategies were tried out.\textsuperscript{18}

The criterion measure, the Wonderlic IQ test, does not appear to be at fault. Previous studies have shown decent construct validation in college populations (e.g., Kennedy et al., 1989; McKelvie, 1994; Wonderlic, 1992). Rather than being inappropriate for measuring IQ in college samples, its lackluster performance corresponds directly to its low standard deviation here. It performed no better and no worse than any standard IQ test would have in this situation.

So why the poor correspondence between the traditional direct and indirect measures of self-rated intelligence and IQ? No association can be observed unless people happen to base their self-perceptions on insight about the same abilities that are tapped by IQ tests. The disynchrony between the other available self-ratings and IQ includes motivated and unmotivated ignorance. The motivated portion involves inflated self-perceptions due to narcissism or self-deception: Previous research shows that this component contributes even more than IQ, perhaps 20 percent of the reliable variance (Gabriel, Critelli, & Ee, 1994; Paulhus, Yik, & Lysy, 1996). This motivated component also includes idiosyncratic definitions of intelligence designed defensively to match their own best abilities and therefore ensure that they are intelligent (Dunning & Cohen, 1992). The unmotivated portion of ignorance may include a lack of interest, concern, or insight into such matters (Campbell & Lavallee, 1993).

Finally, the handicap placed on all of the validities demonstrated in this report must be pointed out. The restriction of range created by the use of college samples is likely to have diminished all validities as a function of the reduced variances (see Cohen & Cohen, 1983). Thus the baseline validity for the single item "intelligent" would have reached .30-.35, instead of .20-.23. Similarly, instead of the cap of .30 for aggregated instruments, values of .40-45 could have been achieved in the normal population. The latter values appear strong enough to be useful in research, if not in diagnosing individuals.

Self-reports of intelligence should not be evaluated solely in terms of potential as proxies for IQ tests. Lay perceivers are
quick to argue that there is more to intelligence than IQ. Participants in the present study, for example, may have based their self-ratings on their creativity, their interpersonal sensitivity, their musical ability, or their self-insight--none of which are tapped by standard IQ tests. A number of expert commentators such as Sternberg and Gardner have argued that scientific conceptions of intelligence more closely to lay conceptions must be tied in.

But what criteria besides IQ could be used to evaluate self-reports of intelligence? Peer-ratings, for one. In another report, evidence that self-report measures predict peer-ratings of intelligence above and beyond IQ has been provided (Lysy & Paulhus, 1996). That is, some portion of observers' perceptions of intelligence is detectable by self and observers but not by IQ tests. The best part of traditional measures is that they correlate with peer-ratings. The OCQ would not measure this "non-test" entity. Thus self-report measures of intelligence have validity beyond their use as proxy IQ measures. From this perspective, it would actually be surprising to find high correlations between IQ tests and perceptions of intelligence.

Performance of the New Indirect Measure: The OCQ

The OCQ has confirmed its promise as an index of general cognitive ability. Because its validity was only evaluated once before, the two replications here provide assurance of its value as a proxy measure of IQ. If the mean validity of .44 is disattenuated for reliability the correlation reaches .61. And, correcting for restriction of range brings the validity all the way up to .76.

It is interesting that the OCQ does not show strong associations with the other self-reports, either direct or indirect. Perhaps the others are more intercorrelated because they are contaminated with self-presentation. Neither the Discrimination index nor the IQ test can contain such contamination. Together, this pattern suggests that the Discrimination index does indeed possess construct validity as an objective measure of IQ. Of course, it may not be as effective as the other measures in predicting broader conceptions of intelligence.
The ability of the OCQ to independently measure discrimination and bias makes it a potentially valuable tool in a number of settings. The OCQ could replace traditional IQ testing in situations where evaluating a candidate's knowledge is cumbersome, for example, when a large number of topics must be addressed, and where administering objective performance measures is difficult. For example, the OCQ can be completed at home or on the telephone rather than under supervised conditions. In addition, the overclaiming notion seems ideal for survey research on such issues as literacy, recognition of political candidates, or recognition of consumer products.

**Other Promising Avenues**

Apart from the OCQ, several potential avenues toward clarifying the links between test performance and self-perceptions of intelligence exist. First is the development of new intelligence tests to encompass more of everyday conceptions of "intelligence". To the extent that test content corresponds to everyday conceptions, then associations should be higher. Wagner and Sternberg (1986) have pursued this avenue by developing objective measures of practical intelligence. Salovey and Mayer's (1990) "Emotional Intelligence" is another measure that shifts the conceptual borders of intelligence toward everyday conceptions.

A second avenue for future research is clarifying and perhaps improving the other side of the relationship, namely, the self-perceptions. What cues are people using to judge their intelligence? The lens model is proving profitable in specifying proximal cues, that is, objective behaviors that trigger attributions of intelligence (Reynolds & Gifford, 1996). Matches and mismatches between self- and peer-perceptions of intelligence and their correlates, too, are being examined (Lysy & Paulhus, 1996). This research should help specify the missing content in current self-report measures.

Finally, further research on the traditional indirect measures studied here is encouraged. Their greatest potential asset has never been directly tested: That is, indirect measures may outperform direct measures in more ego-threatening administration conditions. Another issue worthy of study is whether the direct
ability items work only when interspersed with a variety of other items. Finally, future research should further examine whether some form of objectivity in self-reports is necessary to achieve validity. This appears to be the case given that the objective attribute of the OCQ is one of the most obvious characteristics separating it from the other available measures.

Conclusions
The present report constitutes the most comprehensive examination of self-reports of intelligence to date. The available measures have been organized into four categories of self-rated intelligence to investigate the effects of employing indirect versus direct measures, and the effects of aggregation on predicting objective intelligence. The addition of a new, indirect self-report measure to the potential proxy measures exhausts the available possibilities. Administration of these measures to two large samples leads to a few key conclusions.

1. Both direct and indirect traditional self-report measures of intelligence can reliably predict IQ scores. Because of restricted range of abilities in competitive college samples, however, the validity limit appears to be .30.
2. Direct items about global mental abilities are more valid than indirect items. The one clear exception is the high validities of indirect items referring to enjoyment of reading.
3. Aggregation is beneficial up to a point. With the exception of reading items, aggregation doesn't appear to help beyond 4-5 core items referring directly to close synonyms/antonyms of intelligence (e.g., smart, clever, simple, not gifted).
4. Among available traditional measures, the most effective predictors of IQ were Gough's Ie and Trapnell's Smart scale. Equally effective were Hogan's School Success scale, and Sternberg's Verbal Ability scale.
5. A new indirect measure outperformed all of the available measures. Paulhus' Over-Claiming Questionnaire demonstrated its validity even in a restricted range with a validity cap of about .50.
As a whole, the verdict is pessimistic about the utility of the traditional self-reports as proxy measures of IQ in college samples. The verdict is more optimistic about their utility for assessing intelligence as a broader concept, particularly in the general population. And finally, the verdict is highly optimistic about the potential of Paulhus' Over-Claiming Questionnaire. In any case, researchers who require some proxy scale for their research should benefit from the guidelines that have been provided here.
Footnotes

1. The term, 'validity', is used to mean correlation with a specific criterion. Its use does not imply that IQ is the sole criterion for measuring intelligence.

2. Strictly speaking, Sternberg’s subjects were not all college-educated, but their IQs averaged one standard deviation above the population mean.

3. Note that this instrument is not a checklist in the strict sense of requiring respondents to check off answers dichotomously. Instead, the items are rated as in a Likert format.

4. Despite its conceptual relevance, the item, “ingenious” was not included in Sample 2 because of confusion the item caused. Apparently, some subjects thought the item meant “not a genius”.

5. Two of the Intellect-School Success items are identical to two of the Intellectual efficiency items: They were only included once.

6. A true speeded test comprises all easy questions.

7. The item was “intelligent, clear-thinking” in Study 2.

8. Estimated from the manual norms for men and women weighted according to sex ratio in our samples.

9. Unfortunately, a retest or a parallel form was not administered.

10. The prediction formula when only one measure is lengthened is presented by Thorndike (1982, p.153). Even with an infinite number of equally good items, the prophecy formula would predict an upper limit of .48.

11. These values are likely to be overestimates because of capitalization on chance. Unfortunately, cross-validation from one sample to the other is not feasible because the items were chosen on the basis of consistent performance across both samples.

12. The items were selected by their conceptual similarity to "intelligent" (gifted, smart, clever, etc.).

13. Sternberg(1988) reported using diagnostic weighting of the BCL to improve its validity. With strict adherence to Sternberg’s
method, however, Paulhus, Lysy, and Yik (in press) found no validity improvement.

14.Interestingly, an instrument recently developed to predict school success contains the same two categories of predictors (Giddan, Jurs, Andberg, & Bunnell, 1996).

15.The acceptance of "Sports Illustrated for Kids" in educational software is a good example of this thinking. Although reading such nonintellectual material was traditionally discouraged, the new philosophy is that children are more likely to read if the material focuses on their interests.

16.Although addressing cultural interests, absorption, and imagination, this variable correlates with IQ (McCrae & Costa, 1985).

17.This estimate is based on adjusting the median correlation of the Wonderlic with other IQ tests (.83) for restriction of range (Cohen & Cohen, 1983). Instead of 7.12, the general population SD, the mean SD of our two samples was only 4.6.

18.This is not to imply that aggregation is not useful in psychological assessment. It is most likely the narrowness of the target (an IQ test) that constrained the value of aggregation here.
References


104th meeting of the American Psychological Association, Toronto.


Table 1.
A two-factor categorization of traditional self-report intelligence measures

<table>
<thead>
<tr>
<th>Directness Strategy</th>
<th>Single Item</th>
<th>Aggregated Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct</td>
<td>General rating of intelligence</td>
<td>Composite ratings of intelligence</td>
</tr>
<tr>
<td>Indirect</td>
<td>Individual item from Indirect scales</td>
<td>Full indirect scales <em>(Ie, Smart, Intellect, BCL)</em></td>
</tr>
</tbody>
</table>

Note. Ie refers to Gough's Intellectual efficiency scale; Smart refers to Trapnell's scale; Intellect refers to Hogan's scale; BCL refers to Sternberg's Behavior Check List.
### Table 2.
Descriptive statistics

<table>
<thead>
<tr>
<th>Scale</th>
<th># Items</th>
<th>Scale Min-Max</th>
<th>Item Mean</th>
<th>SD^a</th>
<th>Range^a</th>
<th>Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Intelligence Composite Scale</strong></td>
<td>4</td>
<td>1-5</td>
<td>3.66</td>
<td>.67</td>
<td>3.75</td>
<td>.68</td>
</tr>
<tr>
<td><em>Single-item</em></td>
<td>1</td>
<td>1-5</td>
<td>4.02</td>
<td>.77</td>
<td>4.00</td>
<td>.46^d</td>
</tr>
<tr>
<td><strong>Gough's le scale</strong></td>
<td>52</td>
<td>True-False</td>
<td>.68</td>
<td>--</td>
<td>--</td>
<td>.72</td>
</tr>
<tr>
<td><strong>Trapnell's Smart scale</strong></td>
<td>4</td>
<td>1-9</td>
<td>5.45</td>
<td>1.54</td>
<td>8.00</td>
<td>.86</td>
</tr>
<tr>
<td><strong>Hogan's Intellect scale</strong></td>
<td>22</td>
<td>True-False</td>
<td>.59</td>
<td>--</td>
<td>--</td>
<td>.73</td>
</tr>
<tr>
<td><em>Intellectance</em></td>
<td>15</td>
<td>True-False</td>
<td>.62</td>
<td>--</td>
<td>--</td>
<td>.72</td>
</tr>
<tr>
<td><em>School Success</em></td>
<td>7</td>
<td>True-False</td>
<td>.52</td>
<td>--</td>
<td>--</td>
<td>.61</td>
</tr>
<tr>
<td><strong>Sternberg's BCL</strong></td>
<td>41</td>
<td>1-9</td>
<td>6.24</td>
<td>.82</td>
<td>4.63</td>
<td>.93</td>
</tr>
<tr>
<td><em>Verbal Ability</em></td>
<td>13</td>
<td>1-9</td>
<td>6.02</td>
<td>.96</td>
<td>5.15</td>
<td>.80</td>
</tr>
<tr>
<td><em>Problem Solving</em></td>
<td>15</td>
<td>1-9</td>
<td>6.26</td>
<td>.92</td>
<td>5.87</td>
<td>.89</td>
</tr>
<tr>
<td><em>Social Competence</em></td>
<td>13</td>
<td>1-9</td>
<td>6.45</td>
<td>.84</td>
<td>4.38</td>
<td>.77</td>
</tr>
<tr>
<td><strong>Paulhus' Hits</strong></td>
<td>6^e</td>
<td>0-1</td>
<td>.57^e</td>
<td>--</td>
<td>--</td>
<td>.90</td>
</tr>
<tr>
<td><em>OCQ Discrimination</em></td>
<td></td>
<td></td>
<td>.66</td>
<td>--</td>
<td>--</td>
<td>.89</td>
</tr>
<tr>
<td><em>False Alarms</em></td>
<td>6^e</td>
<td>0-1</td>
<td>.27</td>
<td>--</td>
<td>--</td>
<td>.87</td>
</tr>
</tbody>
</table>

^a Standard Deviation
^b Composite Scale
^c Single-item
^d Alpha
^e Paulhus' Hits
Note. Top row of each cell is from Sample 1 (N = 262 to 279); Bottom row is from Sample 2 (N = 239 to 265) (154 for OCQ Discrimination in Sample 2). Sample size varied due to missing data, i.e., not every subject filled out every measure.

Standard deviations and ranges are calculated across item means rather than across subject means; Since this is not relevant for the two True-False scales these statistics are omitted.

Includes "intelligent" and three conceptually similar items. "intelligent" (Sample 1); "clear-thinking, intelligent" (Sample 2).

Estimate from the mean of the intercorrelations among the four intelligence-related items comprising the direct composite. Statistics presented are the mean across the six categories.

For the OCQ the item mean is the mean proportion (see Results section).
Table 3.
Intercorrelations among indirect scales and subscales

<table>
<thead>
<tr>
<th>Scale</th>
<th>1.</th>
<th>2.</th>
<th>3.</th>
<th>4.</th>
<th>5.</th>
<th>6.</th>
<th>7.</th>
<th>8.</th>
<th>9.</th>
<th>10.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Ie</td>
<td></td>
<td>.08</td>
<td>.47</td>
<td>.32</td>
<td>.43</td>
<td>.41</td>
<td>.36</td>
<td>.39</td>
<td>.29</td>
<td></td>
</tr>
<tr>
<td>2. Smart</td>
<td>.24</td>
<td></td>
<td>.23</td>
<td>.18</td>
<td>.21</td>
<td>.48</td>
<td>.47</td>
<td>.49</td>
<td>.33</td>
<td>.16</td>
</tr>
<tr>
<td>3. Intellect</td>
<td>.27</td>
<td>.29</td>
<td></td>
<td>.90</td>
<td>.65</td>
<td>.56</td>
<td>.58</td>
<td>.51</td>
<td>.43</td>
<td>.31</td>
</tr>
<tr>
<td>4. INT</td>
<td>.27</td>
<td>.24</td>
<td>.92</td>
<td></td>
<td>.25</td>
<td>.51</td>
<td>.50</td>
<td>.49</td>
<td>.38</td>
<td>.22</td>
</tr>
<tr>
<td>5. SS</td>
<td>.24</td>
<td>.25</td>
<td>.68</td>
<td>.35</td>
<td></td>
<td>.36</td>
<td>.42</td>
<td>.27</td>
<td>.29</td>
<td>.30</td>
</tr>
<tr>
<td>6. BCL</td>
<td>.32</td>
<td>.45</td>
<td>.47</td>
<td>.44</td>
<td>.31</td>
<td></td>
<td>.88</td>
<td>.93</td>
<td>.89</td>
<td>.28</td>
</tr>
<tr>
<td>7. VA</td>
<td>.34</td>
<td>.43</td>
<td>.48</td>
<td>.42</td>
<td>.38</td>
<td>.89</td>
<td></td>
<td>.72</td>
<td>.67</td>
<td>.34</td>
</tr>
<tr>
<td>8. PS</td>
<td>.29</td>
<td>.50</td>
<td>.43</td>
<td>.44</td>
<td>.22</td>
<td>.89</td>
<td>.67</td>
<td></td>
<td>.77</td>
<td>.19</td>
</tr>
<tr>
<td>9. SC</td>
<td>.19</td>
<td>.23</td>
<td>.30</td>
<td>.28</td>
<td>.20</td>
<td>.86</td>
<td>.65</td>
<td>.68</td>
<td></td>
<td>.24</td>
</tr>
<tr>
<td>10. OCQ Disc</td>
<td>.34</td>
<td>.10</td>
<td>.10</td>
<td>.07</td>
<td>.11</td>
<td>.19</td>
<td>.26</td>
<td>.11</td>
<td>.13</td>
<td></td>
</tr>
</tbody>
</table>

Note. Correlation coefficients in the upper right of the matrix are from Sample 1 (N = 278); coefficients in the lower left are from Sample 2 (N = 266, except for OCQ Discrimination: N = 137). All correlations above .20 are significant, p < .001, two-tailed. Two identical items from both the Intellectual efficiency and Intellect-School Success scales were given to the Intellect, and subsequently, the Intellect-School Success scale for these calculations.
Table 4.
Correlations of self-report predictors with IQ

<table>
<thead>
<tr>
<th>Scale</th>
<th># of Items</th>
<th>Sample 1</th>
<th>Sample 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Direct Measures</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single-item&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1</td>
<td>.20***</td>
<td>.23***</td>
</tr>
<tr>
<td>Composite scale&lt;sup&gt;b&lt;/sup&gt;</td>
<td>4</td>
<td>.24***</td>
<td>.26***</td>
</tr>
<tr>
<td><strong>Indirect Measures</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gough <em>Ie</em></td>
<td>52</td>
<td>.20***</td>
<td>.34***</td>
</tr>
<tr>
<td>Trapnell <em>Smart</em></td>
<td>4</td>
<td>.24***</td>
<td>.25***</td>
</tr>
<tr>
<td>Hogan <em>Intellect</em></td>
<td>22</td>
<td>.15*</td>
<td>.22***</td>
</tr>
<tr>
<td><em>Intellectance</em></td>
<td>15</td>
<td>.08</td>
<td>.13*</td>
</tr>
<tr>
<td><em>School Success</em></td>
<td>7</td>
<td>.19**</td>
<td>.27***</td>
</tr>
<tr>
<td>Sternberg <em>BCL</em></td>
<td>41</td>
<td>.20***</td>
<td>.13*</td>
</tr>
<tr>
<td><em>Verbal Ability</em></td>
<td>13</td>
<td>.24***</td>
<td>.18**</td>
</tr>
<tr>
<td><em>Problem Solving</em></td>
<td>15</td>
<td>.17**</td>
<td>.10</td>
</tr>
<tr>
<td><em>Social Competence</em></td>
<td>13</td>
<td>.14*</td>
<td>.04</td>
</tr>
<tr>
<td>Paulhus' <em>OCQ-Disc</em></td>
<td>6&lt;sup&gt;c&lt;/sup&gt;</td>
<td>.39***</td>
<td>.49***</td>
</tr>
</tbody>
</table>

**Note.**  
* p < .05; ** p < .01, *** p < .001, two-tailed.
Sample size ranges from 274 to 301 (Sample 1) and from 241 to 265 (Sample 2) (137 for *OCQ Discrimination* with IQ in Sample 2) due to the subject-matching across the three sources of data (i.e., direct, indirect, and IQ).  
<sup>a</sup>"Intelligent" (Sample 1); "clear-thinking, intelligent" (Sample 2).  
<sup>b</sup>Refers to the direct, intelligence ratings including "intelligent" and three conceptually similar items.  
<sup>c</sup>Refers to the 6 categories, each treated as one score, based on each category's 15 items.
Table 5.
Regression of IQ on four traditional indirect predictors

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Sample</th>
<th>Beta</th>
<th>SigF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gough Intellectual efficiency</td>
<td>1</td>
<td>.17</td>
<td>.02</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.29</td>
<td>.00</td>
</tr>
<tr>
<td>Trapnell Smart</td>
<td>1</td>
<td>.22</td>
<td>.00</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.19</td>
<td>.00</td>
</tr>
<tr>
<td>Hogan Intellect</td>
<td>1</td>
<td>.01</td>
<td>.89</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.14</td>
<td>.04</td>
</tr>
<tr>
<td>Sternberg Behavior Check List</td>
<td>1</td>
<td>.02</td>
<td>.84</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>-.12</td>
<td>.11</td>
</tr>
</tbody>
</table>

Note.  N = 275 (Sample 1) and 265 (Sample 2).
Table 6.
Correlations of four types of traditional predictors with IQ

<table>
<thead>
<tr>
<th>Aggregation Strategy</th>
<th>Single Item</th>
<th>Aggregated Items</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sample 1</td>
<td>Sample 2</td>
</tr>
<tr>
<td>Direct Strategy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Direct</td>
<td>.20</td>
<td>.23</td>
</tr>
<tr>
<td></td>
<td>.24</td>
<td>.26</td>
</tr>
<tr>
<td>Indirect Strategy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indirect</td>
<td>.07</td>
<td>.07</td>
</tr>
<tr>
<td></td>
<td>.18</td>
<td>.24</td>
</tr>
</tbody>
</table>

Note. Single-item validities are correlations of IQ with the single item, "intelligent" (Study 1) or "clear-thinking, intelligent" (Study 2); The indirect single-item validities are mean item validities across all 119 items of the four indirect scales. The direct aggregated items' validities are based on the 4-item direct composite scale. The indirect aggregated items' validities are the mean scale validities across all four indirect measures.
Table 7.
Top 10 item validities from the traditional indirect scales

<table>
<thead>
<tr>
<th>Sample 1</th>
<th>Sample 2</th>
<th>Scale</th>
<th>Item Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>.27***</td>
<td>.32***</td>
<td>BCL</td>
<td>reads with high comprehension</td>
</tr>
<tr>
<td>.28***</td>
<td>.26***</td>
<td>BCL</td>
<td>has a good vocabulary</td>
</tr>
<tr>
<td>.27***</td>
<td>.20**</td>
<td>Intellect</td>
<td>As a child I was always reading</td>
</tr>
<tr>
<td>.21***</td>
<td>.21***</td>
<td>Ie/Intellect</td>
<td>I am quite a fast reader</td>
</tr>
<tr>
<td>.18**</td>
<td>.23***</td>
<td>Ie/Intellect</td>
<td>I read at least ten books a year</td>
</tr>
<tr>
<td>.20**</td>
<td>.25***</td>
<td>Smart</td>
<td>Is considered a very &quot;brainy&quot;, scholarly person</td>
</tr>
<tr>
<td>.21***</td>
<td>.21***</td>
<td>Ie</td>
<td>I was a slow learner in school</td>
</tr>
<tr>
<td>.20**</td>
<td>.25***</td>
<td>Smart</td>
<td>considered exceptionally or unusually intelligent</td>
</tr>
<tr>
<td>.20**</td>
<td>.20**</td>
<td>Ie</td>
<td>I seem to be at least as capable and smart as most</td>
</tr>
<tr>
<td>.22***</td>
<td>.17**</td>
<td>Intellect</td>
<td>I would rather read than watch TV</td>
</tr>
</tbody>
</table>

Note. ** p < .01, *** p < .001, two-tailed. N = (275, 265).
Ie refers to Gough's Intellectual efficiency scale; Smart refers to Trapnell's scale; Intellect refers to Hogan's scale; BCL refers to Sternberg's Behavior Check List.
Table 8.
Validities of traditional indirect items: Means within content category

<table>
<thead>
<tr>
<th>Category of Items</th>
<th># Items</th>
<th>Sample 1</th>
<th>Sample 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ability-related</td>
<td>23</td>
<td>.14</td>
<td>.12</td>
</tr>
<tr>
<td>Personality-related</td>
<td>51</td>
<td>.05</td>
<td>.06</td>
</tr>
<tr>
<td>Interest-related</td>
<td>26</td>
<td>.03</td>
<td>.06</td>
</tr>
<tr>
<td>Behavior-related Non-reading</td>
<td>8</td>
<td>.05</td>
<td>.05</td>
</tr>
<tr>
<td></td>
<td>Reading</td>
<td>.19</td>
<td>.18</td>
</tr>
</tbody>
</table>

Note.  \( N = (275, 265) \). Each entry is the mean of all item validities for each category. \(^a\)Number of items refers to the number of item validities used in calculating the mean item validity.
Appendix A
Circle True (T) or False (F) for each statement as it applies to you.

T F 1. I have had very peculiar and strange experiences.

T F 2. It often seems that my life has no meaning.

T F 3. When someone does me a wrong, I feel I should pay them back if I can, just for the principle of the thing.

T F 4. Any job is all right with me, just as long as it pays well.

T F 5. People often expect too much of me.

T F 6. I get sort of annoyed with writers who go out of their way to use strange and unusual words.

T F 7. Success is a matter of will power.

T F 8. When in a group of people I have trouble thinking of the right things to talk about.

T F 9. It is more important that a father be kind than be successful.

T F 10. I was a slow learner in school.

T F 11. I must admit I have no great desire to learn new things.

T F 12. My skin seems to be unusually sensitive to touch.

T F 13. Parents are much too easy on their children nowadays.

T F 14. I read at least ten books a year.
T F 15. I have the wanderlust, and am never happy unless I am roaming or traveling about.

T F 16. My parents have often disapproved of my friends.

T F 17. Teachers often expect too much work from the students.

T F 18. I am not afraid of picking up a disease or germs from door knobs.

T F 19. I am quite a fast reader.

T F 20. I have had more than my share of things to worry about.

T F 21. I feel like giving up quickly when things go wrong.

T F 22. In a group of people I would not be embarrassed to be called upon to start a discussion or give an opinion about something that I know well.

T F 23. I like to read about history.

T F 24. In school, I find it very hard to talk in front of the class.

T F 25. I dread the thought of an earthquake.

T F 26. I like science.

T F 27. I am bothered by people outside, on buses, in stores, etc. watching me.

T F 28. I have never been in trouble because of my sex behavior.

T F 29. I often get feelings like crawling, burning, tingling, or "going to sleep" in different parts of my body.
T F 30. I seem to be at least as capable and smart as most others around me.

T F 31. I like poetry.

T F 32. I am often not in on the gossip of the group I belong to.

T F 33. A windstorm terrifies me.

T F 34. I enjoy a race or game better when I bet on it.

T F 35. Most people make friends because friends are likely to be useful to them.

T F 36. I have never seen a vision.

T F 37. The future seems hopeless to me.

T F 38. I have had no difficulty starting or holding my urine.

T F 39. I often feel as if the world was just passing me by.

T F 40. I get pretty discouraged with the law when a smart lawyer gets a criminal off.

T F 41. The only interesting part of the newspaper is the "funnies" (or the "sports").

T F 42. I gossip a little.

T F 43. At times I have been so entertained by the cleverness of a crook that I have hoped he would not get caught.

T F 44. I have had attacks where I could not control my movements or speech, but where I knew what was going on around me.
T F 45. I have had no difficulty in starting or holding my bowel movement.

T F 46. I like to read about science.

T F 47. I do not read every editorial in the newspaper every day.

T F 48. If people had not had it in for me, I would have been much more successful

T F 49. I daydream very little.

T F 50. I have often been frightened in the middle of the night.

T F 51. I work under a great deal of tension.

T F 52. I seldom worry about my health.
Appendix B
Circle True (T) or False (F) for each statement as it applies to you.

T  F  1. I like classical music.
T  F  2. I enjoy reading poetry.
T  F  3. I find Greek Mythology interesting.
T  F  4. I hate opera singing.
T  F  5. I like to try new, exotic foods.
T  F  6. I would like a job that requires travelling.
T  F  7. I like a lot of variety in my life.
T  F  8. I'm not afraid to be the first to try something.
T  F  9. I like doing things no one else has done.
T  F  10. I enjoy the excitement of the unknown.
T  F  11. I'm known for coming up with good ideas.
T  F  12. I'm not very inventive.
T  F  13. I can make up stories quickly.
T  F  14. I'm good at inventing games, stories and rhymes.
T  F  15. I am a quick-witted person.
T  F  16. I enjoy solving riddles.
T  F  17. I enjoy working crossword puzzles.
T  F  18. I like detective stories.
T  F  19. As a child I was always reading.
T  F  20. I would rather read than watch T.V.
Appendix C
Write a number from 1 to 9 to indicate how much each of the following is true of you. Compare yourself to other UBC students.

Not at all Average Very True
1 2 3 4 5 6 7 8 9

I am the kind of person who...

___ 1. reasons logically and well
___ 2. speaks clearly and articulately
___ 3. accepts others for what they are
___ 4. identifies connections among ideas
___ 5. admits mistakes
___ 6. sees all aspects of a problem
___ 7. converses well
___ 8. displays interest in the world at large
___ 9. keeps an open mind
___ 10. is knowledgeable about (at least) one special field
___ 11. is on time for appointments
___ 12. responds thoughtfully to others' ideas
___ 13. studies hard
___ 14. sizes up situations well
___ 15. reads with high comprehension
___ 16. thinks before speaking and doing
___ 17. gets to the heart of problems
___ 18. reads widely
___ 19. displays curiosity
___ 20. interprets information accurately
21. deals effectively with people
22. does not make snap judgments
23. makes good decisions
24. writes down thoughts easily
25. makes fair judgments
26. goes to original sources for basic information
27. sets aside time for reading
28. is good at assessing the relevance of information to the problem at hand
29. poses problems in an optimal way
30. has a good vocabulary
31. is sensitive to other people's needs and desires
32. is a good source of ideas
33. accepts social norms
34. is frank and honest with self and others
35. perceives implied assumptions and conclusions
36. tries new things
37. displays interest in the immediate environment
38. listens to all sides of an argument
39. deals with problems resourcefully.
40. has a social conscience
41. is verbally fluent
Appendix D
Trapnell Items

Write a number from 1 to 9 to indicate how much each of the following is true of you. Compare yourself to other UBC students.

Not at all  Average  Very True
1    2    3    4    5    6    7    8    9

I am the kind of person who...

___ is considered unusually "gifted" or talented at academic things

___ is considered exceptionally or unusually intelligent

___ is considered a very "brainy", scholarly person

___ usually had grades near the top of every class
Appendix E
FAMILIARITY QUESTIONNAIRE

Instructions

Please rate your familiarity with each item on this questionnaire. Use the scale below as a guide.

<table>
<thead>
<tr>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Never Heard</td>
<td>Very Familiar</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Of It</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For example, if the item said "Bill Clinton" or "Mexico", or "the Bible", you would probably write a '6' beside it because it is very familiar. However, if the item said "Fred Gruneberg" (my next door neighbor) you would write a '0' to indicate you never heard of him.

e.g.,

6 Bill Clinton
0 Fred Gruneberg

In other words, the difficulty of the items ranges from easy to impossible.

We want to determine if individuals who are knowledgeable about one area are also knowledgeable about other areas.
<table>
<thead>
<tr>
<th>Historical Names and Events</th>
<th>Physical Sciences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Napoleon</td>
<td>Manhattan Project</td>
</tr>
<tr>
<td>Robespierre</td>
<td>asteroid</td>
</tr>
<tr>
<td>El Puente</td>
<td>nuclear fusion</td>
</tr>
<tr>
<td>My Lai</td>
<td>cholarine</td>
</tr>
<tr>
<td>The Lusitania</td>
<td>atomic number</td>
</tr>
<tr>
<td>Ronald Reagan</td>
<td>hydroponics</td>
</tr>
<tr>
<td>Prince Lorenzo</td>
<td>alloy</td>
</tr>
<tr>
<td>The Luddites</td>
<td>plate tectonics</td>
</tr>
<tr>
<td>Neville Chamberlain</td>
<td>photon</td>
</tr>
<tr>
<td>Vichy Government</td>
<td>ultra-lipid</td>
</tr>
<tr>
<td>Queen Alberta</td>
<td>centripetal force</td>
</tr>
<tr>
<td>Bay of Pigs</td>
<td>plates of parallax</td>
</tr>
<tr>
<td>Torquemada</td>
<td>nebula</td>
</tr>
<tr>
<td>Wounded Knee</td>
<td>particle accelerator</td>
</tr>
<tr>
<td>Clara Barton</td>
<td>satellite</td>
</tr>
<tr>
<td>Never Heard Of It</td>
<td>Very Familiar</td>
</tr>
<tr>
<td>-------------------</td>
<td>---------------</td>
</tr>
</tbody>
</table>

### Popular Culture Names

- Gail Brennan
- Jackie Robinson
- Houdini
- Ginger Rogers
- Greta Garbo
- Dale Carnegie
- Scott Joplin
- Rube Goldberg
- George Gershwin
- Mae West
- Jesse Owens
- Sister St. Mark
- Louis Lapointe
- King Kong
- P.T. Barnum

### Books and Poems

- Antigone
- Murphy's Last Ride
- Catcher in the Rye
- The Koran
- Hiawatha
- Trapnell's Dilemma
- Mein Kampf
- The Aeneid
- Faustus
- The Boy Who Cried Wolf
- Pygmalion
- Hickory Dickory Dock
- The Divine Comedy
- Windermere Wild
- The Raven
<table>
<thead>
<tr>
<th>Authors and Characters</th>
<th>Social Science and Law</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adonis</td>
<td>yellow journalism</td>
</tr>
<tr>
<td>Mephistopheles</td>
<td>existential angst</td>
</tr>
<tr>
<td>Shylock</td>
<td>nationalism</td>
</tr>
<tr>
<td>Ancient Mariner</td>
<td>megaphrenia</td>
</tr>
<tr>
<td>Doctor Fehr</td>
<td>acrophobia</td>
</tr>
<tr>
<td>Venus</td>
<td>pulse tax</td>
</tr>
<tr>
<td>Romeo and Juliet</td>
<td>pork-barreling</td>
</tr>
<tr>
<td>The Great Graziano</td>
<td>prejudice</td>
</tr>
<tr>
<td>Norman Mailer</td>
<td>Christian Science</td>
</tr>
<tr>
<td>Horatio Alger</td>
<td>ombudsman</td>
</tr>
<tr>
<td>Charlotte Bronte</td>
<td>consumer apparatus</td>
</tr>
<tr>
<td>Artemis</td>
<td>superego</td>
</tr>
<tr>
<td>Lewis Carroll</td>
<td>trust-busting</td>
</tr>
<tr>
<td>Admiral Broughton</td>
<td>behaviorism</td>
</tr>
<tr>
<td>Mrs. Malaprop</td>
<td>Oedipus complex</td>
</tr>
</tbody>
</table>
List of Foils

Historical Names and Events: El Puente, Prince Lorenzo, Queen Alberta

Physical Sciences: cholarine, ultra-lipid, plates of parallax

Popular Culture: Gail Brennan, Sister St. Mark, Louis Lapointe

Books and Poems: Murphy's Last Ride, Trapnells' Dilemma, Windermere Wild

Authors and Characters: Doctor Fehr, The Great Graziano, Admiral Broughton

Social Science and Law: megaphrenia, pulse tax, consumer apparatus
Appendix F
**Over-Claiming Questionnaire mean item familiarity**

<table>
<thead>
<tr>
<th>Category</th>
<th>Mean item familiarity rating</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Existent items</td>
<td>Foils</td>
<td></td>
</tr>
<tr>
<td>Historical Names and Events</td>
<td>1.97</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.12</td>
<td>.54</td>
<td></td>
</tr>
<tr>
<td>Physical Sciences</td>
<td>4.15</td>
<td>1.61</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.16</td>
<td>1.27</td>
<td></td>
</tr>
<tr>
<td>Popular Culture</td>
<td>2.39</td>
<td>.49</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.99</td>
<td>.39</td>
<td></td>
</tr>
<tr>
<td>Books and Poems</td>
<td>2.52</td>
<td>.48</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.83</td>
<td>.26</td>
<td></td>
</tr>
<tr>
<td>Authors and Characters</td>
<td>2.03</td>
<td>.45</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.42</td>
<td>.29</td>
<td></td>
</tr>
<tr>
<td>Social Science and Law</td>
<td>2.97</td>
<td>.68</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.77</td>
<td>.77</td>
<td></td>
</tr>
<tr>
<td>Overall (of all 90 items)</td>
<td>2.67</td>
<td>.79</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.05</td>
<td>.59</td>
<td></td>
</tr>
</tbody>
</table>

**Note.** Top row of each cell is from Sample 1 ($N = 279$); bottom row of each cell is from Sample 2 ($N = 157$).