

**ART**

**ABSTRACT REASONING TEST**

**TECHNICAL MANUAL**



# CONTENTS

---

● Theoretical Overview .....	2
● The Abstract Reasoning Test (ART) .....	4
● The Psychometric Properties of the ART.....	5
● Appendix – Administration Instructions.....	9
● References .....	11

# 1. THEORETICAL OVERVIEW

---

## 1.1 THE ROLE OF INTELLIGENCE TESTS IN PERSONNEL SELECTION AND ASSESSMENT

While much useful information can be gained from the standard job interview, the interview nonetheless suffers from a number of serious weaknesses. Perhaps the most important of these is that the interview has been shown to be a very unreliable way to judge a person's aptitudes and abilities. This is because it is an unstandardised assessment procedure that does not directly assess mental ability, but rather assesses a person's social skills and reported past achievements and performance.

Clearly, the interview provides a useful opportunity to probe each applicant in depth about their work experience, and explore how they present themselves in a formal social setting. Moreover, behavioural interviews can be used to assess an applicant's ability to 'think on their feet' and explain the reasoning behind their decision making processes. Assessment centres can provide further useful information on an applicant by assessing their performance on a variety of work-based simulation exercises. However, interviews and assessment centre exercises do not provide a reliable, standardised way to assess an applicant's ability to solve novel, complex problems that require the use of logic and reasoning ability.

Intelligence tests, on the other hand, do just this; providing a reliable, standardised way to assess an applicant's ability to use logic to solve complex problems. As such, tests of general mental ability are likely to play a significant role in the selection process. Schmidt & Hunter (1998), in their seminal review of the research literature, note that over 85 years of research has clearly demonstrated that general mental ability (intelligence) is the single best predictor of job performance.

From the perspective of assessing a respondent's intelligence, the unstandardised idiosyncratic nature of interviews makes it impossible to directly compare one applicant's ability with another's. Not only do interviews not provide an objective base-line against which to contrast interviewees' differing performances but, moreover, different interviewers typically come to radically different conclusions about the same applicant. Not only do applicants respond

differently to different interviewers asking ostensibly the same questions, but what applicants say is often interpreted quite differently by different interviewers. In such cases we have to ask which interviewer has formed the 'correct' impression of the candidate, and to what extent any given interviewer's evaluation of the candidate reflects the interviewer's preconceptions and prejudices rather than reflecting the candidate's performance.

There are similar limitations on the range and usefulness of the information that can be gained from application forms or CV's. Whilst work experience and qualifications may be prerequisites for certain occupations, in and of themselves they do not predict whether a candidate is likely to perform well or badly in a new position. Moreover, a person's educational and occupational achievements are likely to be limited by the opportunities they have had, and as such may not reflect their true potential. Intelligence tests enable us to avoid many of these problems, not only by providing an objective measure of a person's ability, but also by assessing the person's potential, and not just their achievements to date.

## 1.2 THE ORIGINS OF INTELLIGENCE TESTS

The assessment of mental ability, or intelligence, is one of the oldest areas of research interest in psychology. Gould (1981) has traced attempts to scientifically measure mental acuity, or ability, to the work of Galton in the late 1800s. Prior to Galton's (1869) pioneering research, the assessment of mental ability had focussed on phrenologists' attempts to assess intelligence by measuring the size of people's heads!

Reasoning tests, in their present-day form, were first developed by Binet (1910); a French educationalist who published the first test of mental ability in 1905. Binet was concerned with assessing the intellectual development of children, and to this end he invented the concept of mental age. Questions assessing academic ability were graded in order of difficulty, according to the average age at which children could successfully answer each question. From the child's performance on this test, it was possible to derive the child's mental age. If, for example, a child performed at the level of the average 10

year old on Binet's test then that child was classified as having a mental age of 10, regardless of the child's chronological age.

The concept of the Intelligence Quotient (IQ) was developed by Stern (1912), from Binet's notion of mental age. Stern defined IQ as mental age divided by chronological age multiplied by 100. Previous to Stern's work chronological age had been subtracted from mental age to provide a measure of mental alertness. Stern on the other hand showed that it was more appropriate to take the ratio of these two constructs, to provide a measure of the child's intellectual development that was independent of the child's age. He further proposed that this ratio should be multiplied by 100 for ease of interpretation; thus, avoiding cumbersome decimals.

Binet's early tests were subsequently revised by Terman et al. (1917) to produce the famous Stanford-Binet IQ test. IQ tests were first used for selection by the American military during the First World War, when Yerkes (1921) tested 1.75 million soldiers with the Army Alpha and Army Beta tests. Thus, by the end of the war, the assessment of general mental ability had not only firmly established its place within the discipline of academic psychology, but had also demonstrated its utility for aiding the selection process.

### 1.3 THE CONCEPTS OF FLUID AND CRYSTALLISED INTELLIGENCE

The idea of general mental ability, or general intelligence, was first conceptualised by Spearman in 1904. He reflected on the popular notion that some people are more academically able than others, noting that people who tend to perform well in one intellectual domain (e.g., science) also tend to perform well in other domains (e.g., languages, mathematics, etc.). He concluded that an underlying factor termed general intelligence, or 'g', accounted for this tendency for people to perform well across a range of areas, while differences in a person's specific abilities or aptitudes accounted for their tendency to perform marginally better in one area than in another (e.g., to be marginally better at French than they are at Geography).

Spearman, in his 1904 paper, outlined the theoretical framework underpinning factor analysis; the statistical procedure that is used to identify the shared factor ('g') that accounts for a person's tendency to perform well (or badly) across a range of different tasks. Subsequent developments in the mathematics underpinning

factor analysis, combined with advances in computing, meant that after the Second World War psychologists were able to begin exploring the structure of human mental abilities using these new statistical procedures.

Being most famous for his work on personality, and in particular the development of the 16PF, the pioneering work that Raymond B. Cattell (1967) did on the structure of human intelligence has often been overlooked. Through an extensive research programme, Cattell and his colleagues identified that 'g' (general intelligence) could be decomposed into two highly correlated subtypes of mental ability, which he termed fluid and crystallised intelligence.

Fluid intelligence is reasoning ability in its most abstract and purest form. It is the ability to analyse novel problems, identify the patterns and relationships that underpin these problems and extrapolate from these using logic. This ability is central to all logical problem solving and is crucial for solving scientific, technical and mathematical problems. Fluid intelligence tends to be relatively independent of a person's educational experience and has been shown to be strongly determined by genetic factors. Being the 'purest' form of intelligence, or 'innate mental ability', tests of fluid intelligence are often described as culture fair.

Crystallised intelligence, on the other hand, consists of fluid ability as it is evidenced in culturally valued activities. High levels of crystallised intelligence are evidenced in a person's good level of general knowledge, their extensive vocabulary and their ability to reason using words and numbers. In short, crystallised intelligence is the product of cultural and educational experience in interaction with fluid intelligence. As such it is assessed by traditional tests of verbal and numerical reasoning ability, including critical reasoning tests.

## 2. THE ABSTRACT REASONING TEST (ART)

---

### 2.1 ITEM FORMAT

Crystallised and fluid intelligence are most readily evidenced, respectively, in tests of verbal/numerical and abstract reasoning ability. (See Heim et al. 1974, for examples of such tests). Probably the most famous tests that have been developed on British samples of fluid and crystallised intelligence are; the Raven's Progressive Matrices (RPM) – of which there are three versions (the Coloured Progressive Matrices, The Standard Progressive Matrices and the Advanced Progressive Matrices), and the Mill Hill Vocabulary Scales (MHVS) – of which there are ten versions.

Being measures of crystallised intelligence, tests assessing verbal/numerical reasoning ability (including critical reasoning tests) are strongly influenced by the respondent's educational background. These tests require respondents to draw on their knowledge of words and their meanings, and their knowledge of mathematical operations, to perceive the logical relationships underpinning the tests' items. Tests of abstract reasoning ability, on the other hand, assesses the ability to solve abstract, logical problems which require no prior knowledge or educational experience. As such, these tests are the least affected by the respondent's educational experience; assessing what is often considered to be the purest form of 'intelligence' – or 'innate reasoning ability' – namely fluid intelligence.

To this end, the ART presents respondents with a three-by-three matrix consisting of eight cells containing geometric patterns; the ninth and final cell in the matrix being left blank. Respondents have to deduce the logical rules that govern how the sequence of geometric designs progresses (both horizontally and vertically) across the cells of the matrix, and extrapolate from these rules the next design in the sequence. Requiring little use of language, and no general knowledge, matrix reasoning tests – in the form of the ART or RPM – provide the purest measure of fluid intelligence.

### 2.2 TEST CONSTRUCTION

Research has clearly demonstrated that in order to accurately assess reasoning ability it is necessary to use tests which have been specifically designed to measure the ability being assessed in the population on which the test is intended to be used. This ensures that the test is appropriate for the particular group being assessed. For example, a test designed for those of average ability will not accurately distinguish between people of high ability, as all the respondents' scores will cluster towards the top end of the scale. Similarly, a test designed for people of high ability will be of little practical use if given to people of average ability. Not only will the test not discriminate between respondents, with all their scores clustering towards the bottom of the scale, but also as the questions will mostly be too difficult for the respondents to answer they are likely to lose motivation, thereby further reducing their scores.

The ART was specifically designed to assess the fluid intelligence of technicians, people in scientific, engineering, financial and professional roles, as well as senior managers and those who have to take strategic business decisions. As such the test items were developed for respondents of average, or above average, ability.

The initial item pool was trialled on students enrolled on technical training courses and undergraduate programmes, as well as on a subsample of respondents in full-time employment in a range of professional, managerial and technical occupations. Following extensive trialling, a subset of items that had high levels of internal consistency (corrected item-whole correlations of .4 or greater), and of graded difficulty, were selected for inclusion in the ART.

The test consists of 35 matrix reasoning items, ordered by item difficulty, with the respondent having 30 minutes to complete the test.

## 3 THE PSYCHOMETRIC PROPERTIES OF THE ART

---

### 3.1 STANDARDISATION

Normative data allows us to compare an individual's score on a standardised scale against the typical score obtained from a clearly defined group of respondents (e.g., graduates, the general population, etc.).

To enable any respondent's score on the ART to be meaningfully interpreted, the test was standardised against a population similar to that on which it has been designed to be used (e.g., people in technical, managerial, professional and scientific roles). Such standardisation ensures that the scores obtained from the ART can be interpreted by relating them to a relevant distribution of scores.

### 3.2 RELIABILITY

The reliability of a test assesses the extent to which the variation in test scores is due to true differences between people on the characteristic being measured – in this case fluid intelligence – or to random measurement error.

Reliability is generally assessed using one of two different methods; one assesses the stability of the test's scores over time, the other assesses the internal consistency, or homogeneity, of the test's items.

#### 3.2.1 Reliability: Temporal Stability

Also known as test-retest reliability, this method for assessing a test's reliability involves determining the extent to which a group of people obtain similar scores on the test when it is administered at two points in time. In the case of reasoning tests, where the ability being assessed does not change substantially over time (unlike personality), the two occasions when the test is administered may be many months apart. If the test were perfectly reliable, that is to say test scores were not influenced by any random error, then respondents would obtain the same score on each occasion, as their level of intelligence would not have changed between the two points in time when they completed the test. In this way, the extent to which respondents' scores are unstable over time can be used to estimate the test's reliability.

Stability coefficients provide an important indicator of a test's likely usefulness. If these coefficients are low, then this suggests that the test is not a reliable measure and is therefore of little practical use for assessment and selection purposes.

#### 3.2.2 Reliability: Internal Consistency

Also known as item homogeneity, this method for assessing a test's reliability involves determining the extent to which, if people score well on one item, they also score well on the other test items. If each of the test's items were a perfect measure of intelligence, that is to say the score the person obtained on the items was not influenced by any random error, then the only factor that would determine whether a person was able to answer each item correctly would be the item's difficulty. As a result, each person would be expected to answer all the easier test items correctly, up until the point at which the items became too difficult for them to answer. In this way, the extent to which respondents' scores on each item are correlated with their scores on the other test items, can be used to estimate the test's reliability.

The most commonly used internal consistency measure of reliability is Cronbach's (1960) alpha coefficient. If the items on a scale have high intercorrelations with each other, then the test is said to have a high level of internal consistency (reliability) and the alpha coefficient will be high. Thus, a high alpha coefficient indicates that the test's items are all measuring the same thing, and are not greatly influenced by random measurement error. A low alpha coefficient on the other hand suggests that either the scale's items are measuring different attributes, or that the test's scores are affected by significant random error. If the alpha coefficient is low this indicates that the test is not a reliable measure, and is therefore of little practical use for assessment and selection purposes.

### 3.3 VALIDITY

The fact that a test is reliable only means that the test is consistently measuring a construct, it does not indicate what construct the test is consistently measuring. The concept of validity addresses this issue. As Kline (1993) notes 'a test is said to be valid if it measures what it claims to measure'.

An important point to note is that a test's reliability sets an upper bound for its validity. That is to say, a test cannot be more valid than it is reliable because if it is not consistently measuring a construct it cannot be consistently measuring the construct it was developed to assess. Therefore, when evaluating the psychometric properties of a test its reliability is usually assessed before addressing the question of its validity.

There are two principal ways in which a test can be said to be valid.

#### 3.3.1 Validity: Construct Validity

Construct validity assesses whether the characteristic which a test is measuring is psychologically meaningful and consistent with how that construct is defined. Typically, the construct validity of a test is assessed by demonstrating that the test's results correlate with other major tests which measure similar constructs and do not correlate with tests that measure different constructs. (This is sometimes referred to as a test's convergent and discriminant validity). Thus, demonstrating that a test which measures fluid intelligence is more strongly correlated with an alternative measure of fluid intelligence than it is with a measure of crystallised intelligence, would be evidence of the measure's construct validity.

#### 3.3.2 Validity: Criterion Validity

This method for assessing the validity of a test involves demonstrating that the test meaningfully predicts some real-world criterion. For example, a valid test of fluid intelligence would be expected to predict academic performance, particularly in science and mathematics.

Moreover, there are two types of criterion validity - predictive validity and concurrent validity. Predictive validity assesses whether a test is capable of predicting an agreed criterion which will be available at some future time, e.g., can a test of fluid intelligence predict

future GCSE maths results. Concurrent validity assesses whether the scores on a test can be used to predict a criterion which is available at the time the test was completed, e.g., can a test of fluid intelligence predict a scientist's current publication record.

### 3.4 ART: STANDARDISATION AND BIAS

#### 3.4.1 Standardisation Sample

The ART was standardised on a sample of 651 adults of working age, drawn from a variety of managerial, professional and graduate occupations. The mean age of the standardisation sample was 30.4 years, with 32% of the sample being women. 24% of the sample identified themselves as being of non-white (European) ethnic origin. Of these respondents 22% identified themselves as being of Black African ethnic origin, 12% of Indian origin, 9% of Black Caribbean origin and 8% of Pakistani origin.

#### 3.4.2 Gender and Age Effects

Gender differences on the ART were examined by comparing the scores that a group of male and a group of female graduate level bankers obtained on the ART. These data are presented in Table 1 and clearly indicate no significant sex difference in ART scores, suggesting that this test is not sex biased.

### 3.5 ART: RELIABILITY

#### 3.5.1 Reliability: Internal Consistency

Table 2 presents alpha coefficients for the ART on a number of different samples. Inspection of this table indicates that all these coefficients are above .8, indicating that the ART has good levels of internal consistency reliability.

#### 3.5.2 Reliability: Test-retest

As noted above, test-retest reliability estimates the test's reliability by assessing the temporal stability of the test's scores. As such, test-retest reliability provides an alternative measure of reliability to internal consistency estimates of reliability, such as the alpha coefficient. Theoretically, test-retest and internal consistency estimates of a test's reliability



should be equivalent to each other. The only difference between the test-retest reliability statistic and the alpha coefficient is that the latter provides an estimate of the lower bound

of the test's reliability and, as such, the alpha coefficient for any given test is usually lower than the test-retest reliability statistic.

Table 1. Mean and standard deviation in parentheses for ART scores for male (n=146) and female (n=63) bankers and the associated t statistic

Mean and Standard Deviation		t-value	df	p
Males	Females			
21.31 (4.35)	21.15 (4.56)	.022	.207	0.82

Table 2. Alpha coefficients for different samples on the ART

Sample	Sample size	Sample size
Graduate level bankers	n=209	.81
Retail Managers	n=105	.81
Undergraduates	n=176	.85

### 3.6 ART: VALIDITY

#### 3.6.1 The Relationship Between the ART and the RAPM

The Ravens Advanced Progressive Matrices (RAPM) is one of the most well respected, and best validated, measures of fluid intelligence. A sample of 32 undergraduates completed the ART and RAPM for research purposes. The substantial correlation between these two measures  $.77$  ( $p < .001$ ) was highly significant. This large correlation therefore provides strong support for the construct validity of the ART, indicating that these two measures are assessing related constructs.

#### 3.6.2 The relationship between the ART and the GRT1 numerical reasoning test

The GRT1 numerical reasoning test is a measure of crystallised intelligence that has been specifically designed for graduate populations. 209 graduate level bankers completed this test along with the ART. These two tests were found to be significantly correlated ( $r = .49$ ,  $p < .001$ ) with each other, providing good support for the construct validity of the ART.

#### 3.6.3 The relationship between the ART and the CRTB

The Critical Reasoning Test Battery (CRTB) assesses verbal and numerical critical reasoning; the ability to make logical inferences on the basis of, respectively, written text or numerical data presented in tables and graphs. Both verbal and numerical critical reasoning ability are different types of crystallised intelligence, and would therefore be expected to be modestly correlated with the ART, which measures fluid intelligence. A sample of 272 applicants for managerial positions with a large multinational company completed the CRTB and ART as part of a selection processes. The ART was found to correlate significantly with both the Verbal ( $r = .32$ ,  $p < .001$ ) and Numerical ( $r = .38$ ,  $p < .001$ ) Critical Reasoning tests, providing support for the construct validity of the ART.

#### 3.6.4 The relationship between the ART and learning style

A sample of 143 undergraduates completed the ART and the Learning Styles Inventory (LSI) – a measure of learning style. As would

be predicted, fluid intelligence was found to be modestly correlated ( $r=.29$ ,  $p<.001$ ) with a more abstract rather than a more concrete learning style, and with a more holistic ( $r=.23$ ,  $p<.001$ ) rather than a more serial (i.e., focussing on the 'big picture' rather than fine details) learning style, providing further support for the construct validity of the ART.

### 3.6.5 The relationship between the ART and Intellectance

Intellectance is a meta-cognitive variable that assesses a person's perception of their general mental ability. While it is a personality factor, rather than an ability factor, intellectance has nonetheless consistently been found to correlate with objective assessments of mental ability. As such it would be expected to be modestly correlated with fluid intelligence. A sample of 132 applicants for managerial and professional posts completed the 15FQ+ and the ART as part of an assessment process. The ART was

found to be correlated ( $r=.32$ ,  $p<.001$ ) with the 15FQ+ Factor Intellectance ( $\beta$ ), providing further support for the construct validity of the ART.

### 3.6.6 Mean differences in ART scores by occupational group

Table 3 presents the mean scores (and the associated 95% confidence intervals) on the ART obtained by three occupational groups of differing ability levels; retail managers, bankers and graduate engineers. Given the nature of their work, engineers would be expected to have the highest level of fluid intelligence followed by bankers then retail managers. Inspection of Table 3 indicates that these means conform to this pattern, with the 95% confidence intervals indicating that these differences are unlikely to be due to chance effects. These data therefore provide further support for the construct validity of the ART.

Table 3. Mean ART scores for each occupational group and associated 95% confidence intervals

Sample	Sample Size	Mean and 95% Confidence Interval
Graduate Engineers	n=40	24.03 $\pm$ 3.3
Graduate Bankers	n=209	21.20 $\pm$ 1.2
Retail Managers	n=105	15.2 $\pm$ 2.2

## APPENDIX – ADMINISTRATION INSTRUCTIONS

### BEFORE STARTING THE QUESTIONNAIRE

Put candidates at their ease by giving information about yourself: the purpose of the test; the timetable for the day; whether or not the questionnaire is being completed as part of a wider assessment programme, and how the results will be used and who will have access to them. Ensure that you and other administrators have requested that all mobile phones have been switched off, etc. The instructions below should be read out verbatim and the same script should be followed each time the ART is administered to one or more candidates. Instructions for the administrator are printed in ordinary type. Instructions designed to be read aloud to candidates have lines marked above and below them, are in bold and enclosed by speech marks. If this is the first or only questionnaire being administered, give an introduction as per or similar to the following example:

*" From now on, please do not talk amongst yourselves, but ask me if anything is not clear. Please ensure that any mobile telephones, pagers or other potential distractions are switched off completely. We shall be doing the Abstract Reasoning Test which is timed for 30 minutes. During the test I shall be checking to make sure you are not making any accidental mistakes when filling in the answer sheet. I will not be checking your responses to see if you are answering correctly or not. "*

**WARNING:** It is most important that answer sheets do not go astray. They should be counted out at the beginning of the test and counted in again at the end.

### DISTRIBUTE THE ANSWER SHEETS

Then ask:

*" Has everyone got two sharp pencils, an eraser, some rough paper and an answer sheet. "*

Rectify any omissions, then say:

*" Print your last name, first name clearly on the lines provided. Indicate your preferred title by checking the title box; then note your gender and age. Please insert today's date which is [ ] in the Date of Testing box. "*

If biographical information is required, ask respondents to complete the biodata section. If answer sheets are to be scanned, explain and demonstrate how the ovals are to be completed, emphasising the importance of fully blackening the oval.

Walk around the room to check that the instructions are being followed.

**WARNING:** It is vitally important that test booklets do not go astray. They should be counted out at the beginning of the session and counted in again at the end.

### DISTRIBUTE THE BOOKLETS WITH THE INSTRUCTION:

*" Please do not open the booklet until instructed to do so. "*

Remembering to read slowly and clearly, go to the front of the group and say:

*" Please open the booklet at Page 2 and follow the instructions for this test as I read them aloud. "*

Pause to allow booklets to be opened.

*" This is a test of your ability to perceive the relationships between abstract shapes and figures. The test consists of 35 questions and you will be having 30 minutes in which to attempt them. "*

*Before you start the timed test, there are two example questions on Page 3 to show you how to complete the test. For each question you will be presented with a 3 by 3 array of boxes, the last of which is empty. Your task is to select which of the six possible answers presented below fit vertically and horizontally to complete the sequence. One and only one answer is correct in each case.*

"

Check for understanding of the instructions so far, then say:

*" You now have a chance to complete the example questions in order to make sure that you understand the test. Please attempt the example questions now.*

"

While the candidates are doing the examples, walk around the room to check that everyone is clear about how to fill in the answer sheet. Make sure that nobody looks through the actual test items after completing the example questions. When everyone has finished the example questions (allow a maximum of one and half minutes), give the answers as follows:

*" The answer to Example 1 is number 3 and the answer to Example 2 is also number 3, as only these answers complete the sequence.*

"

Then say:

*" Before you begin the timed test, please note the following points:  
 1 Attempt each question in turn. If you are unsure of an answer, mark your best choice but avoid wild guessing.  
 2 If you wish to change an answer, simply erase it fully before marking your new choice of answer.  
 3 Time is short, so do not worry if you do not finish the test in the available time.  
 4 Work quickly, but not so quickly as to make careless mistakes.*

*1 You have 35 questions and 30 minutes in which to complete them.*

*If you reach the End of Test before time is called, you may review your answers if you wish.*

"

Then say very clearly:

*" Is everybody clear about how to do this test?*

"

Deal with any questions appropriately, then starting a stop-watch or setting a count-down timer on the word

**BEGIN**, say:

*" Please turn over the page and begin.*

"

Answer only questions relating to procedure at this stage, but enter in the Administrator's Test Record any other problems which occur. Walk around the room at appropriate intervals to check for potential problems. At the end of the 30 minutes say clearly:

*" STOP NOW please and close your booklet.*

"

You should intervene if candidates continue beyond this point.

Collect answer sheets and test booklets, ensuring that all materials are returned (count booklets and answer sheets)

Then say:

*" Thank you for completing the Abstract Reasoning Test.*

"

## REFERENCES

---

- Binet, A. (1910) *Les idées modernes sur les enfants*. Paris: E. Flammarion.
- Cattell, R. B. (1967). The theory of fluid and crystallised intelligence. *British Journal of Educational Psychology*, 37, 209-224.
- Cronbach, L.J. (1960). *Essentials of Psychological Testing (2nd Edition)*. New York: Harper.
- Galton F. (1869). *Hereditary Genius*. London: MacMillan.
- Gould, S.J. (1981). *The Mismeasure of Man*. Harmondsworth, Middlesex: Pelican.
- Heim, A.H., Watt, K.P. and Simmonds, V. (1974). *AH2/AH3 Group Tests of General Reasoning; Manual*. Windsor: NFER Nelson.
- Kline, P. (1993). *Personality: The Psychometric View*. London: Routledge.
- Schmidt, F.L., & Hunter, J.E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124, 262-274.
- Spearman, C. (1904). General-intelligence; objectively defined determined and measured. *American Journal of Psychology*, 15, 210-292.
- Stern, W. (1912). *Psychologische Methoden der Intelligenz-Prüfung*. Leipzig, Germany: Earth.
- Terman, L.M. et. al., (1917). *The Stanford Revision of the Binet-Simon scale for measuring intelligence*. Baltimore: Warwick and York.
- Yerkes, R.M. (1921). *Psychological examining in the United States army*. *Memoirs of the National Academy of Sciences*, 15.