# CRTB2

CRITICAL REASONING
TEST BATTERY
technical manual

# CONTENTS

②

# LIST OF TABLES

④

# 1

# THEORETICAL OVERVIEW

# THE ROLE OF PSYCHOMETRIC TESTS IN PERSONNEL SELECTION AND ASSESSMENT

A major reason for using psychometric tests to aid selection decisions is that they provide information that cannot be obtained easily in other ways. If such tests are not used then what we know about the applicant is limited to the information that can be gleaned from an application form or CV, an interview and references. If we wish to gain information about a person's specific aptitudes & abilities and about their personality, attitudes and values then we have little option but to use psychometric tests. In fact, psychometric tests can do more than simply provide additional information about the applicant. They can add a degree of reliability and validity to the selection procedure that it is impossible to achieve in any other way. How they do this is best addressed by examining the limitations of the information obtained through interviews, application forms and references and exploring how some of these limitations can be overcome by using psychometric tests.

While much useful information can be gained from the interview, which clearly has an important role in any selection procedure, it does nonetheless suffer from a variety of weaknesses. Perhaps the most important of these is that the interview as been shown to be a very unreliable way to judge a person's character. This is because it is an unstandardised assessment procedure. That is to say, each interview will be different from the last. This is true even if the interviewer is attempting to ask the same questions and act in the same way with each applicant. It is precisely this aspect of the interview that is both its main strength and its main weakness. The interview enables us to probe each applicant in depth and discover individual strengths and weaknesses. Unfortunately, the interviews unstandardised, idiosyncratic nature makes it difficult to compare applicants, as it provides no base line against which to contrast interviewees' differing performances. In addition, it is likely that different interviewers may come to radically different conclusions about the same applicant. Applicants will respond differently to different interviewers, quite often saying very different things to them. In addition, what any one applicant might say will be interpreted quite differently by each interviewer. In such cases we have to ask which interviewer has formed the correct impression of the candidate? This is a question to which there is no simple answer.

A further limitation of the interview is that it only assesses the candidate's behaviour in one setting, and with regard to a small number of people. How the candidate might act in different situations and with different people (e.g. when dealing with people on the shop floor) is not assessed, and cannot be predicted from an applicant's interview performance. Moreover, the interview provides no reliable information about the candidate's aptitudes and abilities. The most we can do is ask the candidate about his strengths and weaknesses, a procedure that has obvious limitations. Thus the range, and reliability of the informa-

tion that can be gained through an interview are limited.

There are similar limitations on the range and usefulness of the information that can be gained from application forms or CV's. While work experience and qualifications may be prerequisites for certain occupations, in and of themselves they do not determine whether a person is likely to perform well or badly. Experience and academic achievement is not always a good predictor of ability or future success. While such information is important it may not be sufficient on its own to enable us to confidently choose between applicants. Thus aptitude and ability tests are likely to play a significant role in the selection process as they provide information on a person's potential and not just their achievements to date. Moreover, application forms tell us little about a person's character. It is often a candidate's personality that will make the difference between an average and an outstanding performance. This is particularly true when candidates have relatively similar records of achievement and past performance. Therefore, personality tests can play a major role in assisting selection decisions.

There is very little to be said concerning the usefulness of references. While past performance is undoubtedly a good predictor of future performance references are often not good predictors of past performance. If the name of the referee is supplied by the applicant, then it is likely that they have chosen someone they expect to speak highly

of them. They will probably have avoided supplying the names of those who may have a less positive view of their abilities. Aptitude and ability tests, on the other hand, give us an indication of the applicant's probable performance under exam conditions. This is likely to be a true reflection of the person's ability.

What advantages do psychometric tests have over other forms of assessment? The first advantage they have is that they add a degree of reliability to the selection procedure that cannot be achieved without their use. Test results can be represented numerically making it easy both to compare applicants with each other, and with pre-defined groups (e.g. successful vs. unsuccessful job incumbents). In the case of personality tests the test addresses the issue of how the person characteristically behaves in a wide range of different situations and with different people. Thus psychometric tests, both personality tests and aptitude and ability tests provide a range of information that are not easily and reliably assessed in other ways. Such information can fill important gaps which have not been assessed by application forms, interviews and references. It can also raise questions that can later be directly addressed in the interview. It is for this reason that psychometric tests are being used increasingly in personnel selection. Their use adds a degree of breadth to assessment decisions which cannot be achieved in any other way.

## RELIABILITY AND VALIDITY

As previously noted, besides providing information that cannot be easily obtained in other ways psychometric tests also add reliability and validity to the selection procedure. There are two ways in which psychometric tests increase the reliability of the assessment procedure:

### i) The use of a standardised assessment procedure:

Reliability is achieved by using the same tests on each applicant and administering, scoring and interpreting the test results in the same way. Thus, individual biases and distortions are removed from the assessment procedure. By comparing each applicant's scores against an agreed norm we create a baseline that enables us not only to compare applicants with each other, but also to contrast them against some agreed criterion (e.g. against the performance of a sample of graduates, accountants etc.). Thus, subjective and idiosyncratic interpretations of a candidate's performance are removed from the assessment process.

### ii) The use of well standardised & reliable psychometric tests:

To ensure the assessment procedure produces reliable and consistent results it is necessary to use well-constructed psychometric tests. It is not sufficient simply to administer any questionnaire that purports to be a psychometric test, or assessment system. If the test has been constructed badly, it will neither be reliable nor valid and will add little to the assessment process. In the most extreme case the use of such a test may invalidate an otherwise valid selection procedure. For a test to be reliable each of the questions in each scale must be a good measure of the underlying trait that the scale is attempting to assess. To this end the test publisher should provide data to demonstrate that the test is both reliable and valid. (The statistics that are used to determine this are described later in the manual).

# THE ORIGINS OF REASONING TESTS

The assessment of intelligence or reasoning ability is perhaps one of the oldest areas of research interest in psychology. Gould (1981) has traced attempts to scientifically measure psychological aptitudes and abilities to the work of Galton at the end of the last century. Prior to Galton's pioneering work, however, interest in this area was aroused by phrenologists' attempts to assess mental ability by measuring the size of people's heads. Reasoning tests, in their present form, were first developed by Binet, 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 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 complete each item. From the child's performance on such a test it was possible to derive its mental age. This involved comparing the performance of the child with the performance of the 'average child' from different age groups. If the child performed at the level of the average 10 year old, then the child was said to have a mental age of 10, regardless of its chronological age. From this idea the concept of the Intelligence Quotient (IQ) was developed by William Stern (1912) who defined it as mental age divided by chronological age multiplied by 100. Previous to Stern's paper chronological age had been subtracted from mental age to provide a measure of mental alertness. Stern showed that it was more appropriate to take the ratio of these two constructs, which would provide a measure of the child's intellectual development relative to other children. 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 now famous Stanford-Binet IQ test. These early IQ tests were first used for selection by the American's during the first world war, when Yerkes (1921) tested 1.75 million soldiers with the army alpha and beta tests. Thus by the end of the war, the assessment of reasoning ability had firmly established its place within psychology.

# 2

# THE CRITICAL
# REASONING TESTS

# THE DEVELOPMENT OF THE CRITICAL REASONING TESTS

Research has clearly demonstrated that in order to accurately assess reasoning ability it is necessary to develop tests which have been specifically designed to measure that ability in the population under consideration. That is to say, we need to be sure that the test has been developed for use on the particular group being tested, and thus is appropriate for that particular group. There are two ways in which this is important. Firstly, it is important that the test has been developed in the country in which it is intended to be used. This ensures that the items in the test are drawn from a common, shared cultural experience, giving each candidate an equal opportunity to understand the logic which underlies each item. Secondly, it is important that the test is designed for the particular ability range on which it is to be used. A test designed for those of average ability will not accurately distinguish between people of high ability as all the scores will cluster towards the top end of the scale. Similarly, a test designed for people of high ability will be of little use if given to people of average ability. Not only will it not discriminate between applicants, as all the scores will cluster towards the bottom of the scale, but also as the questions will be too difficult for most of the applicants they are likely to be de-motivated, producing artificially low scores. Consequently, the VCR2 and NCR2 have been developed on data from undergraduates. That is to say, people of above average intelligence, who are likely to find themselves in senior management positions as their career develops.

In constructing the items in the VCR2 and NCR2 a number of guide lines were borne in mind. Firstly, and perhaps most importantly, special care was taken when writing the items to ensure that in order to correctly solve each item it was necessary to draw logical conclusions and inferences from the stem passage/table. This was done to ensure that the test was assessing critical (logical/deductive) reasoning rather than simple verbal/numerical checking ability. That is to say, the items assess a person's ability to think in a rational, critical way and make logical inferences from verbal and numerical information, rather than simply check for factual errors and inconsistencies.

In order to achieve this goal for the Verbal Critical Reasoning (VCR2) test two further points were born in mind when constructing the stem passages for the VCR2. Firstly, the passages were kept fairly short and cumbersome grammatical constructions were avoided, so that a person's scores on the test would not be too affected by reading speed; thus providing a purer measure of critical reasoning ability. Secondly, care was taken to make sure that the passages did not contain any information which was counter-intuitive, and was thus likely to create confusion.

To increase the acceptability of the test to applicants the themes of the stem passages were chosen to be relevant to a wide range of business situations. As a consequence of these constraints the final stem passages were similar in many ways to the short articles found in the financial pages of a daily newspaper, or trade magazines.

The second edition of the Verbal and Numerical Critical Reasoning tests has been revised To meet the following aims:

- To impove the face validity of the test items, thus increasing the test's acceptability to respondents.
- To modernise the items to reflect contemporary business and financial issues.
- To improve the tests' reliability and validity while maintaining the tests' brevity – with the CRBT being administrable in under one hour.
- To simplify test scoring.
- To make available a hand scored as well as a computer scored version of the tests.
- To remove the impact of guessing on raw VCR2scores, thus increasing the power of the VCR2 to discriminate between respondents.

As noted above the most significant change in the second edition of the VCR2 has been the incorporation of a correction for guessing. This obviates the problem that, due to the three-point response scale that is used in most verbal critical reasoning test, it is possible for respondents to get 33% of the items correct simply by guessing.

While a variety of methods have been proposed for solving this problem (including the use of negative or harsh scoring criteria) we believe that a correction for guessing is the most elegant and practical solution to this problem.

This correction is based on the number of items the respondent gets wrong on the test. We know that to get these items wrong the respondent must have incorrectly guessed the answer to that item. We can further assume that, by chance, the respondent incorrectly guessed the answer 66% of the time and correctly guessed the answer 33% of the time. Thus it is possible to estimate the number of correct guesses the respondent made from the number of incorrect responses. This correction can then be subtracted from the total score to adjust for the number of items the respondent is likely to have correctly guessed.

The use of this correction improves the test's score distribution, increasing its power to discriminate between the respondents' 'true' ability level. Thus it is recommended that test users correct sores for guessing before standardising scores.

However, as the norm tables for corrected and uncorrected scores are significantly different from each other it is **important**, if hand scoring the Critical Reasoning tests, to ensure that the correct norm table is used to standardise the scores on the VCR2. That is to say, either the norm table for the uncorrected (Appendix IV - Table 2) or corrected scores (Appendix IV - Table 3) depending upon whether or not the correction for guessing has been applied).

# 3

# THE PSYCHOMETRIC PROPERTIES OF THE CRITICAL REASONING TESTS

*This chapter presents information describing the psychometric properties of the Verbal and Numerical Critical Reasoning tests. The aim will be to show that these measures meet the necessary technical requirements with regard to standardisation, reliability and validity, to ensure the psychometric soundness of these test materials.*

# INTRODUCTION

## STANDARDISATION – NORMATIVE

Formative data allows us to compare an individuals score on a standardised scale against the typical score obtained from a clearly identifiable, homogeneous group of people.

## RELIABILITY RELIABILITY

The property of a measurement which assesses the extent to which variation in measurement is due to true differences between people on the trait being measure or to measurement error. In order to provide meaningful interpretations, the reasoning tests were standardised against a number of relevant groups. The constituent samples are fully described in the next section. Standardisation ensures that the measurements obtained from a test can be meaningfully interpreted in the context of a relevant distribution of scores. Another important technical requirement for a psychometrically sound test is that the measurements obtained from that test should be reliable. Reliability is generally assessed using two specific measures, one related to the stability of scale scores over time, the other concerned with the internal consistency, or homogeneity of the constituent items that form a scale score.

## RELIABILITY – ASSESSING STABILITY

Also known as test-retest reliability, an assessment is made of the similarity of scores on a particular scale over two or more test occasions. The occasions may be from a few hours, days, months or years apart. Normally Pearson correlation coefficients are used to quantify the simi-larity between the scale scores over the two or more occasions. Stability coefficients provide an important indicator of a test's likely usefulness of measurement. If these coefficients are low (< approx. 0.6) then it is suggestive of either that the abilities/behaviours/attitudes being measured are volatile or situationally specific, or that over the duration of the retest interval, situational events have made the content of the scale irrelevant or obsolete. Of course, the duration of the retest interval provides some clue as to which effect may be causing the unreliability of measurement. However, the second measure of a scales reliability also provides valuable information as to why a scale may have a low stability coefficient. The most common measure of internal consistency is Cronbach's Alpha. If the items on a scale have high inter-correlations with each other, and with the total scale score, then coefficient alpha will be high. Thus a high coefficient alpha indicates that the items on the scale are measuring very much the same thing, while a low alpha would be suggestive of either scale items measuring different attributes or the presence of error.

## RELIABILITY – ASSESSING INTERNAL CONSISTENCY

Also known as scale homogeneity, an assessment is made of the ability of the items in a scale to measure the same construct or trait. That is a parameter can be computed that indexes how well the items in a scale contribute to the overall measurement denoted by the scale score. A scale is said to be internally consistent if all the constituent item responses are shown to be positively associated with their scale score. The fact that a test has high internal consistency & stability coefficients only guarantees that it is measuring something consistently. It provides no guarantee that the test is actually measuring what it purports to measure, nor that the test will prove useful in a particular situation. Questions concerning what a test actually measures and its relevance in a particular situation are dealt with by looking at the tests validity. Reliability is generally investigated before validity as the reliability of test places an upper limit on tests validity. It can be mathematically demonstrated that a validity coefficient for a particular test can not exceed that tests reliability coefficient.

## VALIDITY

The ability of a scale score to reflect what that scale is intended to measure. Kline's (1993) definition is 'A test is said to be valid if it measures what it claims to measure'. Validation studies of a test investigate the soundness and relevance of a proposed interpretation of that test. Two key areas of validation are known as criterion validity and construct validity.

## VALIDITY – ASSESSING CRITERION VALIDITY

Criterion validity involves translating a score on a particular test into a prediction concerning what could be expected if another variable was observed. The criterion validity of a test is provided by demonstrating that scores on the test relate in some meaningful way with an external criterion. Criterion validity comes in two forms – predictive & concurrent. 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 predict the likelihood of someone successfully completing a training course. Concurrent validity assesses whether the scores on a test can be used to predict a criterion measure which is available at the time of the test – e.g. can a test predict current job performance.

## VALIDITY – ASSESSING CONSTRUCT VALIDITY

Construct validity assesses whether the characteristic which a test is actually measuring is psychologically meaningful and consistent with the tests definition. The construct validity of a test is assessed by demonstrating that the scores from the test are consistent with those from other major tests which measure similar constructs and are dissimilar to scores on tests which measure different constructs.

# STANDARDISATION

The critical reasoning tests were standardised on a mixed sample of 365 people drawn from graduate, managerial and professional groups. The age and sex breakdowns of the normative sample for the VCR2 and NCR2 are presented in Tables 1 and 2 respectively. As would be expected from an undergraduate sample the age distribution is skewed to the younger end of the age range of the general population. The sex distribution is however broadly consistent with that found in the general population.

Norm tables for the VCR2 and NCR2 are presented in Appendix IV. For the Verbal Critical Reasoning test different norm tables are presented for test scores that have, or have not, been corrected for guessing. (A correction for guessing has not been made available for the Numerical Critical Reasoning test as the six-point scale this test uses mitigates against the problem of guessing.) As noted above it is recommended that scores on the VCR2 are corrected for guessing. The correction for guessing should be applied to the raw score (i.e. to the score before it has been standardised.) The corrected (or uncorrected) raw score is then standarised with reference to the appropriate norm table (Appendix IV Table 2 for uncorrected scores and Table 3 for corrected scores.) **Thus it is important that particular care is taken to refer to the correct norm table when standradising VCR2 raw scores.**

In addition, for users of the GeneSys system normative data is available also from within the software, which computes for any given raw score, the appropriate standardised scores for the selected reference group. In addition the GeneSys™ software allows users to establish their own in-house norms to allow more focused comparison with profiles of specific groups.

# BIAS

## GENDER AND AGE DIFFERENCES

Gender differences on CRTB were examined by comparing samples of males and female respondents matched, for educational and socio-economic status. Table 2 opposite provides mean scores for men and women on the verbal and numerical critical reasoning tests, along with the F-ratio for the difference between theses means. While the men in this sample obtained marginally higher scores on both the verbal and numerical reasoning tests, this was not statistically significant.

| Age<br>Mean | Age<br>SD | Male | Female |
|---|---|---|---|
| 31.7 | 7.9 | n=245 | n=119 |

*Table 1 – Mean and SD of age, and gender breakdown, of the normative sample*

| | mean | | F-ratio | Significance<br>of difference |
|---|---|---|---|---|
| | men<br>(n=218) | women<br>(n=166) | | |
| **VCR2** | 21.1 | 22.1 | .64 | n.s. |
| **NCR2** | 9.0 | 10.1 | .15 | n.s. |

*Table 2 – Mean scores for men and women (MBAs) on the VCR2 and NCR2*

# RELIABILITY OF THE CRITICAL REASONING TESTS

If a reasoning test is to be used for selection and assessment purposes the test needs to measure each of the aptitude or ability dimensions it is attempting to measure reliably, for the given population (e.g. graduate entrants, senior managers etc.). That is to say, the test needs to be consistently measuring each ability so that if the test were to be used repeatedly on the same candidate it would produce similar results. It is generally recognised that reasoning tests are more reliable than personality tests and for this reason high standards of reliability are usually expected from such tests. While many personality tests are considered to have acceptable levels of reliability if they have reliability coefficients in excess of .7, reasoning tests should have reliability coefficients in excess of .8.

## GRT2 INTERNAL CONSISTENCY

Table 3 presents alpha coefficients for the Verbal and Numerical Critical Reasoning tests. Each of these reliability coefficients is substantially greater than .8, clearly demonstrating that the VCR2 and NCR2 are highly reliable across a range of samples.

| alpha | Insurance Sales Agents (n=132) | MBA's (n=205) | Undergraduates (n=70) |
|---|---|---|---|
| VCR2 | .88 | .84 | .88 |
| NCR2 | .83 | .81 | .86 |

*Table 3 – Alpha coefficients for the Verbal and Numerical Critical Reasoning Tests*

# VALIDITY

Whereas reliability assess the degree of measurement error of a reasoning test, that is to say the extent to which the test is consistently measuring one underling ability or aptitude, validity addresses the question of whether or not the scale is measuring the characteristic it was developed to measure. This is clearly of key importance when using a reasoning test for assessment and selection purposes. In order for the test to be a useful aid to selection we need to know that the results are reliable and that the test is measuring the aptitude it is supposed to be measuring. Thus after we have examined a test's reliability we need to address the issue of validity. We traditionally examine the reliability of a test before we explore its validity as reliability sets the lower bound of a scale's validity. That is to say a test cannot be more valid than it is reliable.

# STRUCTURE OF THE CRITICAL REASONING TESTS

Specifically we are concerned that the tests are correlated with each other in a meaningful way. For example, we would expect the Verbal and Numerical Critical Reasoning tests be moderately correlated with each other as they are measuring different facets of critical reasoning ability – namely verbal and numerical ability. Thus if the VCR2 and NCR2 were not correlated with each other we might wonder whether each is a good measure of critical reason-ing ability. Moreover, we would expect the Verbal and Numerical Critical Reasoning tests Not to be so highly correlated with each other as to suggest that they are measuring the same construct (i.e. we would expect the VCR2 and NCR2 to show discriminant validity). Consequently, the first way in which we might assess the validity of a reasoning test is by exploring the relationship between the tests.

## THE GRADUATE REASONING TESTS (GRT1) THE GENERAL REASONINGTHE CRITICAL REASONING

Table 4, which presents the Pearson Product moment correlation between the VCR2 and NCR2, demonstrates that while the Verbal and Numerical tests are significantly correlated, they are nevertheless measuring distinct abilities.

| Insurance Sales Agents (n=132) | MBA's (n=170) | Undergraduates (n=70) |
|---|---|---|
| .40 | .57 | .49 |

*Table 4 – Correlations between the VCR2 and NCR2*

# CONSTRUCT VALIDITY OF THE CRITICAL REASONING TESTS

As an evaluation of construct validity, the Verbal and Numerical Critical Reasoning tests were correlated with other widely used measures of related constructs.

The VCR2 and NCR2 were correlated with the APIL-B (Ability, Processing of Information and Learning Battery) that has been developed by Taylor (1995). The APIL-B has been specifically developed to be a culture fair assessment tool for use in a multi-racial context (South Africa). As such, it has been designed to assess an individual's core cognitive capabilities, rather than specific skills that may depend upon educational experience and life advantagement/disadvantagement.

Table 5 presents the correlations between the Verbal and Numerical Critical Reasoning tests with the APIL-B, on a sample of MBA students. These correlations are highly statistically significant, and substantial in size, providing strong support for the construct validity of the VCR2 and NCR2.

The VCR2 and NCR2 were also found to correlate substantially (r=.42 and r=.36 respectively) with Factor B (Intellectual Self-confidence) on the 16PFi on a sample (n=132) of insurance sales agents. This suggests that those respondents who were more confident of their own intellectual ability had higher levels of critical reasoning ability; providing some tangential support for the construct validity of the VCR2 and NCR2.

Table 6 presents the correlations between the original edition of the Verbal and Numerical Critical Reasoning tests and the AH5 – a widely respected measure of general reasoning ability. These data thus provide evidence demonstrating that the first edition of these two tests measure reasoning ability rather than some other (related) construct (i.e. verbal or numerical checking ability). As was noted above, because of the nature of critical reasoning tests items, it is particularly important when developing such tests to demonstrate that they are measuring reasoning ability, and not checking ability. This is demonstrated by inspection of table 6.

The relationship between the first edition of the CRTB and the Watson-Glaser Critical Thinking Appraisal was examined by Correlating the VCR2 and NCR2 with the W-GCTA. The correlations with the W-GCTA were .38, for both the Verbal and Numerical tests. While modest, these correlations nonetheless demonstrate a degree of congruence between these two tests, as would be expected from different measures of critical reasoning.

| | APIL-B | sample size | significance |
|---|---|---|---|
| **VCR2** | .569 | n=250 | p<.001 |
| **NCR2** | .512 | n=169 | p< .001 |

*Table 5 – Correlations between the Verbal and Numerical Critical Reasoning tests with the APIL-B*

| | **VCR2** | **NCR2** |
|---|---|---|
| Verbal/Numerical subscale of the AH5 | .60 | .51 |

*Table 6 – Correlations between the original versions of the VCR2 and NCR2 with the AH5*

# CRITERION VALIDITY OF THE CRITICAL REASONING TESTS

In this section, we provide details of a number of studies in which the critical reasoning tests have been used to predict job related performance criteria.

## INSURANCE SALES

A sample of 132 Insurance Sales Agents completed the CRTB as part of a validation study. The association between their scores on the VCR2 and NCR2 and their job performance was examined using t-tests. Job incumbents were classified as either successful or unsuccessful depending upon their performance after one year in post. Table 7 presents the mean scores for these two groups on the VCR2 and NCR2. Inspection of this table indicates that, on average, the successful incumbents had significantly higher scores on these tests than did the non-successful incumbents. The difference in scores between these two groups reached statistical significance for the NCR2. This provides strong support for the criterion-related validity of this test.

A group of MBA students completed the VCR2 and NCR2 prior to enrolling. Their scores on these tests were then correlated with their performance across different courses on the MBA syllabus. The results of this analysis are presented in Table 8. Inspection of table 8 indicates that the critical reasoning tests were predictive of performance across a number of areas of study. This provides strong support for the predictive validity of the CRTB.

|  | Mean (n=29) | Mean (n=23) | t-value | p |
|---|---|---|---|---|
|  | unsuccessful | successful |  |  |
| **VCR2** | 18.13793 | 21.21739 | 1.47715 | n.s. |
| **NCR2** | 9.72414 | 12.60870 | 2.18352 | < .05 |

*Table 7 – Association between the VCR2, NCR2 and insurance sales success*

| | VCR2 | NCR2 |
|---|---|---|
| Innovation & design | .374 (n=89, p< 01) | .260 (n=89 p< .01) |
| Business decision making | .467 (n=35, p<.01) | .433 (n=35 p<.01) |
| Macro economics | .478 (n=89, p<.001) | .386 (n=89, p<.001) |
| IT | .468 (n=35, p<.01) | .511 (n=35 p<.01) |
| Post Graduate Diploma in Business Administration Average to date | .364 (n=34, p<.05) | .510 (n=34, p<.01) |
| Economics | .236 (n=56, n.s.) | .013 (n=56, n.s.) |
| Analytical Tools and Techniques | .312 (n=51, p<.05) | .134 (n=51, n.s.) |
| Marketing | .204 (n=53, n.s.) | -.124 (n=53, n.s.) |
| Finance & Accounting | .209 (n=56, n.s.) | -.007 (n=56, n.s.) |
| Organisational Behaviour | .296 (n=56, p<.05) | -.032 (n=56, n.s.) |
| MBA Category | .389 (n=48, p<.01) | .109 (n=48, n.s.) |

*Table 8 – Correlations between the VCR2, NCR2 and MBA performance*

# 4 REFERENCES

Binet. A (1910) *Les idées modernes sur les enfants* Paris: E. Flammarion.

Cronbach L.J. (1960) *Essentials of Psychological Testing (2nd Edition)* New York: Harper.

Galton F. (1869) *Heriditary Genuis* London: MacMillan.

Gould, S.J. (1981). *The Mismeasure of Man.* Harmondsworth, Middlesex: Pelican.

Heim, A.H. (1970). *Intelligence and Personality.* Harmondsworth, Middlesex: Penguin.

Heim, A.H., Watt, K.P. and Simmonds, V. (1974). *AH2/AH3 Group Tests of General Reasoning; Manual.* Windsor: NFER Nelson.

Jackson D.N. (1987) *User's Manual for the Multidimensional Aptitude Battery* London, Ontario: Research Psychologists Press.

Johnson, C., Blinkhorn, S., Wood, R. and Hall, J. (1989). *Modern Occupational Skills Tests: User's Guide.* Windsor: NFER-Nelson.

Budd R.J. (1991) *Manual for the Clerical Test Battery*: Letchworth, Herts UK: Psytech International Limited

Budd R.J. (1993) *Manual for the Technical Test Battery:* Letchworth, Herts UK: Psytech International Limited

Stern W (1912) *Psychologische Methoden der Intelligenz-Prüfung.* Leipzig, Germany

Barth Terman, L.M. et. al., (1917). *The Stanford Revision of the Binet-Simon scale for measuring intelligence.* Baltimore: Warwick and York

Watson & Glaser (1980) *Manual for the Watson-Glaser Critical Thinking Appraisal* Harcourt Brace Jovanovich: New York

Yerkes, R.M. (1921). *Psychological examining in the United States army.* Memoirs of the National Academy of Sciences, 15.

# 5

APPENDIX I

## ADMINISTRATION INSTRUCTIONS

Good practice in test administration requires the assessor to set the scene before the formal administration of the tests. This scene-setting generally includes: welcome and introductions; the nature, purpose and use of the assessment and feedback arrangements.

If only one (either the Verbal or Numerical) of the Critical Reasoning tests is being administered then Say:

'From now on, please do not talk among yourselves, but ask me if anything is not clear. If you have a mobile phone please ensure that it is switched off. We shall be doing only one of the two tests contained in the booklet that I will shortly be distributing.

Say **either:**

The Verbal Critical Reasoning Test which takes 15 minutes

**or**

The Numerical Critical Reasoning Test which takes 25 minutes

Continue

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.'

If you are administering both the Verbal and Numerical Critical Reasoning tests (as is more common), and if this is the first or only questionnaire being administered give an introduction as per or similar to the example script provided.

Continue by using the instructions exactly as given. Say:

> 'From now on, please do not talk among yourselves, but ask me if anything is not clear. If you have a mobile phone please ensure that it is switched off. We shall be doing two tests, the Verbal Critical Reasoning Test which takes 15 minutes and the Numerical Critical Reasoning Test which takes 25 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. Please note the answer boxes are in columns (indicate) and remember do not write on the booklets.'

Rectify any omissions, then say:

> 'Print your last name and first name on the line provided, and indicate your title and sex followed by your age and today's date.'

Explain to the respondents what to enter in the boxes marked 'Test Centre' and 'Comments'. Walk round 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.'

Remembering to read slowly and clearly, go to the front of the group. If you are only administering the Numerical Critical Reasoning test then go the section below head Numerical Critical Reasoning test. If you are administering both Critical Reasoning tests, or if you are just administering the Verbal Critical Reasoning test 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).

'In this test you have to draw inferences from short passages of text. You will be presented with a passage of text followed by a number of statements. Your task is to decide, on the basis of the information contained in the passage, whether each statement is true, false or cannot be inferred from the passage. Your decision should be based only on the information contained in the passage and not on your own knowledge or opinions.'

'Mark your answer by filling in the appropriate box, on your answer sheet, that corresponds to your choice.'

'You now have a chance to complete the example questions on page 3 in order to make sure that you understand the test. Enter your responses to the example questions in the section marked Example Questions at the top of the answer sheet.'

Point to the section on the answer sheet marked Example Questions (as you read the above).

Then pause while candidates read the instructions, then say:√

'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 no-one is looking at the actual test items during the example session. When all have finished (allow a maximum of two and a half minutes) give the answers as follows:

'The correct response to Example 1 is **False**. It is explicitly stated within the text that further growth in the number of radio stations is limited due to there being no new radio frequencies available.'

'The correct response to Example 2 is **True**. It is explicitly stated that audience figures affect advertising revenue, thus affecting profitability.'

'The correct response to Example 3 is **Cannot Determine**. It is impossible to infer, from the information provided in the text, whether radio stations in general will become more profitable. The text indicates that audience figures arecurrently poor for many radio stations and that it is expected that some may go bankrupt. However, it is not possible to infer from this that audience figures (and as a result advertising revenue) will increase for the remaining radio stations.'

Check for understanding, then say:

> 'Time is short so when you begin the timed test work as quickly and as accurately as you can.
>
> If you are unsure of answer, mark your best choice and move on to the next question.
>
> If you want to change an answer cross it out, as indicated in the instructions in the top left-hand corner of the answer sheet, and fill in your new choice of answer.'

Point to the top left-hand corner of the answer sheet.

Then continue:

> 'There are 8 passages of text and a total of 40 questions. You have 15 minutes in which to answer the questions.
>
> If you reach the **'End of Test'** before time is called you may review your answers if you wish.
>
> If you have any questions please ask now, as you will not be able to ask questions once the test has started.'

Then say very clearly:

> 'Is everyone clear about how to do this test?'

Deal with any questions, appropriately, then, starting stop watch or setting acount-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 15 minutes, say:

> 'Stop'

You should intervene if candidates continue after this point.

If you are only administering the Verbal Critical Reasoning test say:

> 'Close the test booklets'

**COLLECT ANSWER SHEETS AND BOOKLETS, ENSURING THAT ALL MATERIALS ARE RETURNED (COUNT BOOKLETS AND ANSWER SHEETS)**

Then say:

'Thank you for completing the Critical Reasoning Test Battery'

If you are administering both of the Critical Reasoning tests continue by saying:

' Now please turn to Page 12 which is a blank page'

Then say:

'We are now ready to start the n ext test. Has everyone still-got two sharpened pencils, an eraser, some unused rough paper?'

If not, rectify, then say:

'The next test follows on the same answer sheet, please locate the section now.'

Check for understanding.

Then say:

'Now please turn to page 14...'

If you are only administering the Numerical Critical Reasoning test say:

'Please open the booklet at Page 14...'

and continue by saying:

'and follow the instructions for this test as I read them aloud.' (Pause to allow booklets to be opened).

In this test you will have to draw inferences from numerical information which is presented in tabular form.

You will be presented with a numerical table and asked a number of questions about this information. You will then have to select the correct answer to each question from one of six possible choices. One and only one answer is correct in each case.

Mark your answer, by filling in the appropriate box, on your answer sheet that corresponds to your choice.

You now have a chance to complete the example questions on Pages 15 in order to make sure that you understand the test. Enter your responses to the example questions in the section marked Example Questions at the top of the answer sheet.'

Point to the section on the answer sheet marked Example Questions (as youread the above).

Then pause while candidates read the instructions, then say:

'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 no-one is looking at the actual test items during the example session. When all have finished (allow a maximum of three minutes) give the answers as follows:

'The correct answer to Example 1 is Design (answer no. 5). It can be seen, in the table, that amongst women, design was consistently chosen by the lowest percentage as the most important feature of a car.

The correct answer to Example 2 is performance (answer no. 1). It can be seen that of all the features of a car, perfor-mance is rated by men as being the most important featureof a car.

The correct answer to Example 3 is 10.4 (answer no.5). Of men below the age of 30, 5% identified safety and 52% identified performance as the most important feature of a car. 52 over 5 is 10.4, therefore the answer is number 5.

Please do not turn over the page yet'

Then say:

'Time is short so when you begin the timed test work as quickly and as accurately as you can.

If you want to change an answer cross it out, as indicated in the instructions in the top left-hand corner of the answer sheet, and fill in your new choice of answer.'

Point to the top left-hand corner of the answer sheet.

Then continue:

'There are 6 tables of numerical information and a total of 25 questions. You have 25 minutes in which to answer the questions.

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

If you have any questions please ask now, as you will not be able to ask questions once the test has started.'

Then say very clearly:

'Is everyone clear about how to do this test?'

Deal with any questions, appropriately, then, starting stop watch or setting acount-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 25 minutes, say:

'Stop. Close the test booklets'

You should intervene if candidates continue after this point.

If you are only administering the Verbal Critical Reasoning test say:

**COLLECT ANSWER SHEETS AND BOOKLETS, ENSURING THAT ALL MATERIALS ARE RETURNED (COUNT BOOKLETS AND ANSWER SHEETS)**

Then say either:

'Thank you for completing the Critical Reasoning Test Battery'

or

'Thank you for completing the Numerical Critical Reasoning Test'

# APPENDIX II
# SCORING INSTRUCTIONS

The completed answer sheets are scored and profiled by following the steps listed below:

**1** Remove the top cover sheet of the combined answer/scoring sheet to reveal the scoring key.

To score and standardise the VCR2 follow steps 2-8. To score and standardise the NCR2 follow steps 9-10.

**2** Count up the number of correct responses for the VCR2 and enter the total in the box marked 'Total' (Raw Score).

If you do not wish to correct the VCR2 score for guessing go straight to step 7.

**3** To correct the VCR2 score for guessing add up the total number of incorrect responses (i.e. the total number of items attempted minus the raw score) and enter this in the box marked 'Number Wrong'.

**4** The correction for guessing can be found in Appendix III. The number of incorrect responses is listed in the first column of this table and the corresponding correction for guessing is listed in the second column. Make note of the correction for guessing (that corresponds to the number of incorrectly completed items).

**5** To obtain the corrected raw score, subtract the correction for guessing from the raw score. If this number is negative (i.e. the number corrected for guessing is larger than the raw score) then the corrected raw score is zero. Enter the corrected raw score in the box marked 'Corrected/Uncorrected Raw Score'. To indicate that you have made the correction, delete 'Uncorrected'.

**6** To standardise the corrected raw score, look this up in the norm table presented in Appendix IV – Table 3 and enter this in the box marked 'Standard Score'.

You have scored and standardised the VCR2. If you wish to score and standardise the NCR2 follow steps 9-10.

**7** Enter the total score obtained from step 2 in the box marked 'Corrected/Uncorrected Raw Score'. To indicate that you have not made the correction, delete 'Corrected'.

**8** To standardise the uncorrected raw score, look this value up in the norm table presented in Appendix IV – Table 2 and enter this in the box marked 'Standard Score'.

**9** Count up the number of correct responses to the NCR2 and enter the total in the box marked 'Total'.

**10** To standardise the raw score, look this value up in the norm table presented in Appendix IV – Table 1 and enter this in the box marked 'Standard Score'.

# APPENDIX III
# CORRECTION FOR
# GUESSING

| Number of incorrect answers | Correction *(to be deducted from raw score)* |
|---|---|
| 1 | .5 |
| 2 | 1 |
| 3 | 1.5 |
| 4 | 2 |
| 5 | 2.5 |
| 6 | 3 |
| 7 | 3.5 |
| 8 | 4 |
| 9 | 4.5 |
| 10 | 5 |
| 11 | 5.5 |
| 12 | 6 |
| 13 | 6.5 |
| 14 | 7 |
| 15 | 7.5 |
| 16 | 8 |
| 17 | 8.5 |
| 18 | 9 |
| 19 | 9.5 |
| 20 | 10 |
| 21 | 10.5 |
| 22 | 11 |
| 23 | 11.5 |
| 24 | 12 |
| 25 | 12.5 |
| 26 | 13 |
| 27 | |
| 28 | |
| 29 | |
| 30 | |
| 31 | |
| 32 | Corrected |
| 33 | Raw Score = 0 |
| 34 | |
| 35 | |
| 36 | |
| 37 | |
| 38 | |
| 39 | |
| 40 | |

# APPENDIX IV
# NORM TABLES

| Sten Values | NCR2 Raw |
|---|---|
| 1 | 0-2 |
| 2 | 3 |
| 3 | 4-5 |
| 4 | 6-7 |
| 5 | 8-10 |
| 6 | 11-13 |
| 7 | 14-16 |
| 8 | 17-18 |
| 9 | 19-20 |
| 10 | 21-25 |

*Table 1 – Norm: NCR2 Graduates/ Managers*

| Sten Values | VCR2 Raw |
|---|---|
| 1 | 0-7 |
| 2 | 8-10 |
| 3 | 11-12 |
| 4 | 13-16 |
| 5 | 17-20 |
| 6 | 21-23 |
| 7 | 24-27 |
| 8 | 28-29 |
| 9 | 30-32 |
| 10 | 33-40 |

*Table 2 – Norm: VCR2 (Uncorrected) Graduates/Managers*

| Sten Values | VCR2 Correct |
|---|---|
| Data not yet available | |

*Table 3 – Norm: VCR2 Corrected Graduates/ Managers*