# W-GCTA$^{UK}$ and RANRA

## User Guide


PEARSON

**Published by**

Pearson
Assessment
80 Strand
London WC2R 0RL

**T:** 0845 630 888

**F:** 0845 630 5555

**E:** info@psychcorp.co.uk

www.psychcorp.co.uk

ISBN 978 0 749107 71 0


Copyright © 2008 Pearson Education Ltd or its affiliate(s). All rights reserved.

## Contents



PEARSON

PEARSON

# 1. Purpose of the W-GCTA$^{UK}$ and RANRA User Guide

The following guide is a straightforward and practical handbook for the Watson-Glaser Critical Thinking Appraisal-UK edition (W-GCTA$^{UK}$) and Rust Advanced Numerical Reasoning Appraisal (RANRA) for trained test users.

This guide provides all of the practical information required to use the tests in a work setting, including the application of the tests, a step-by-step guide to administering and scoring the tests, and necessary technical information.

This guide is designed to complement the Test Manuals. Those wishing to gain a greater understanding of the research, development and technical aspects of these tests should refer to the Test Manuals.

PEARSON

## 2. Application of the Tests

The Watson Glaser Critical Thinking Appraisal-UK edition (W-GCTA$^{UK}$) and Rust Advanced Numerical Reasoning Appraisal (RANRA) are psychometric tests of critical thinking and numerical reasoning. They measure skills relevant to problem solving and decision-making in a variety of graduate and managerial roles. Critical thinking can be defined as the ability to identify and analyse problems, and seek and evaluate relevant information in order to reach appropriate conclusions. Numerical reasoning, as assessed by RANRA, focuses on estimation and analysis of numerical data.

The tests are suitable for use in a variety of organisational contexts, including selection, development and career counselling across commercial, industrial and public sector organisations. Reference groups are available for the UK population, management roles, and a variety of other positions.

The W-GCTA$^{UK}$ and RANRA may be used separately or together to provide greater breadth of assessment of critical thinking and numerical reasoning. With a timed administration session it takes around 1 hour and 40 minutes to administer both tests (50 minutes and 30 minutes respectively plus administration time)[1].

The W-GCTA$^{UK}$ and RANRA are available in paper and pencil, on-screen and online formats. Both tests and all formats require supervised and controlled administration by a qualified Test Administrator. Those controlling the tests must hold the British Psychological Society Level A Certificate of Competence in Occupational Testing[2]. The tests can be scored by hand or using the Pearson Assessment Bureau Service. Computer generated reports are available for the test taker, examples of these can be found in appendix E.

### 2.1 General description of the W-GCTA$^{UK}$ and RANRA

The W-GCTA$^{UK}$ and RANRA are tests of power. That is, they measure the quality and depth of critical reasoning facility rather than the speed at which a person can perform. They can be administered timed or untimed[1]. The items are of varying difficulty, using multiple choice formats, and are scored as right or wrong. It is highly unlikely that a test taker will answer all items correctly or incorrectly.

[1] Timed administration is recommended as the generous time limits are designed so that 90 percent of test takers will complete the tests within the alloted time. There are instances where administration may be reduced or untimed. Users should refer to section 3.4 of the W-GCTA$^{UK}$ and RANRA manuals.

[2] Refer to the Psychological Testing Centre website for details (www.psychtesting.co.uk)

### 2.1.1 W-GCTA$^{UK}$

Items: 80

Format: Multiple-choice

Timings:

| | |
|---|---|
| Preparation: | 5-10 minutes |
| Administration: | 50 minutes time[1]. |
| Scoring: | 5 minutes (where not automated) |
| Feedback: | 5-10 minutes for verbal or written feedback (expert reports are available from the Pearson Assessment Bureau Service, see appendix E). |

The W-GCTA$^{UK}$ consists of five subtests requiring different, though interdependent, applications of analytical reasoning.

The material in the subtests reflects information typical of that found in the media that requires comment and should not to be accepted unquestioningly or without a degree of critical evaluation. A series of propositions follow each statement and the test taker must evaluate these using a multiple-choice format.

*Test 1* ***Inference***

Discriminating among degrees of truth or falsity of inferences drawn from given information.

*Test 2* ***Recognition of Assumptions***

Identifying unstated assumptions or presuppositions in given statements.

*Test 3* ***Deduction***

Determining whether certain conclusions follow from the information in given statements or premises.

*Test 4* ***Interpretation***

Weighing evidence and deciding if generalisations or conclusions based on data are warranted.

*Test 5* ***Evaluation of Arguments***

Distinguishing between arguments that are strong and relevant and those that are weak or irrelevant to a particular question or issue.

The subtest scores provide useful diagnostic information in development and guidance contexts, but should not be used in recruitment and selection (see section 2.2 of this guide).

[1] Refer to section 3.4 in the W-GCTA$^{UK}$ manual for details on reduced or untimed administration.

### 2.1.2 RANRA

Items: 32

Format: Multiple-choice

Timings:

| | |
|---|---|
| Preparation: | 5-10 minutes |
| Administration: | 30 minutes timed[1]. |
| Scoring: | 5 minutes (where not automated) |
| Feedback: | 5-10 minutes for verbal or written feedback (expert reports are available from the Pearson Assessment Bureau Service, see appendix E). |

The RANRA consists of two subtests that require the understanding and application of numerical reasoning. The material allows test takers to reflect on solutions and evaluate whether or not they make sense. The test has a multiple-choice format.

*Test 1* ***Comparison of Quantities***

Evaluating two quantities and deciding whether they are equal or identifying which one is greater than the other.

*Test 2* ***Sufficiency of Information***

Determining whether the two sets of numerical information provided are sufficient to allow the question to be answered.

The subtest scores provide useful diagnostic information in a development and counselling context, but should not be used in recruitment and selection (see section 2.2 of this guide).

## 2.2 Contexts

The W-GCTA$^{UK}$ and RANRA can be used in employment contexts to assist with selection, development and career guidance and outplacement. The use of these tests in any context should be considered by an individual holding the British Psychological Society Level A Certificate of Competence in Occupational Testing.

[1] Refer to section 3.4 in the RANRA manual for details of a reduced or untimed adminstration.

### 2.2.1 Selection

The W-GCTA$^{UK}$ and RANRA can be used to predict success in jobs that require critical thinking skills.

The assessment of aptitude for analysing and interpreting information, problem solving and decision making is a critical part of effective selection processes for high level (graduate and managerial) roles.

Tests of reasoning ability have been shown to be the most effective single predictor of job performance and training success (e.g. Salgado *et al*., 2004; Robertson & Smith, 2001; Schmidt & Hunter, 1998; 2004). This means that by using reasoning tests, such as the W-GCTA$^{UK}$ and RANRA, you can make more informed decisions on an applicant's ability to do a job and succeed in it. This leads to a reduction in poor recruitment decisions.

Results from the W-GCTA$^{UK}$ and RANRA may be used as an initial screen or, in combination with other assessment methods to provide a full profile of an applicant.

The W-GCTA$^{UK}$ and RANRA each provide a single score which represents a broad measurement across critical thinking and numerical reasoning domains respectively. These scores should be used in combination with other assessment techniques.

Before using the W-GCTA$^{UK}$ and RANRA as part of the selection process organisations should ensure that the tests are relevant to the role. Using inappropriate tests, or relevant tests in an inappropriate manner, can result in poor and unfair decisions. Job analysis and validation of the tests in the relevant context should be carried out.

If you require assistance on validation of the W-GCTA$^{UK}$ and RANRA contact Pearson Assessment.

#### Sales Selection

There are two applicants for the role of sales manager. Sharon, who had a lively and persuasive manner, impressed the selectors at interview with her broad perspective. Martin responded well to the questions but was a little hesitant. The test results showed that while both had adequate numerical reasoning skills (50th and 60th percentile on RANRA respectively); Sharon's critical reasoning was quite weak (30th percentile), whereas Martin was well above average compared to other sales manager applicants at the 70th percentile. The selectors agreed that problem solving and decision making were key elements of the role and with the test results there was no doubt that Martin was the better candidate overall.

### 2.2.2 Development

Psychometric tests can be helpful in better understanding a person's strengths and weaknesses so that appropriate development goals and activities can be set.

The W-GCTA$^{UK}$ and RANRA allow both a broad and in-depth analysis of a person's critical thinking and numerical reasoning skills. W-GCTA$^{UK}$ and RANRA scores can be broken down by subtest to allow a full exploration of the strengths and weaknesses within these skills. It is beyond the scope of this User Guide to advise on development plans for critical thinking or numerical reasoning skills. However, the following references may be useful:

| | |
|---|---|
| Critical Thinking Community: | http://www.criticalthinking.org |
| Edward de Bono: | www.edwarddebono.com |
| Books: | Thomson: *Critical Reasoning, A Practical Introduction* (Routledge) |
| | Fisher: *Critical Thinking: An Introduction* (Cambridge) |
| | Butterworth & Thwaites: *Thinking Skills* (Cambridge University Press) |

Awareness of their critical thinking ability allows people to better understand their own strengths and weaknesses. They can then consider ways to build on their strengths and minimise the impact of their weaknesses. This might be through appropriate career choices and work strategies or identifying development opportunities where it is possible to work on areas which are weaker. There are some training courses which help people develop critical thinking skills. Awareness of limitations allows a person to take actions which will mitigate their impact. Someone who is weak in an area of critical thinking might, for instance, develop the habit of discussing important or complex decisions with a colleague who is stronger in this area before taking action.

It is recommended that these tests be used with other assessments to create a clear and complete picture of an individual, for example personality inventories or 360 appraisals. Tests over- or under-interpreted or used in isolation can lead to poor advice being given as the 'whole person' is not taken into account.

#### Development Environment

Jasmine has been working for the environment office of her local authority since she completed her degree three years ago. She participated in a programme run by her employer to help her develop her career. As part of this she took the W-GCTA$^{UK}$ and RANRA. Her profile showed that while her overall W-GCTA$^{UK}$ score was good (65th percentile) she was much weaker on the 'Recognition of Assumptions' and 'Interpretation' subtests. Her mentor suggested they discuss some of the data and information she used on a day-to-day basis to identify assumptions and interpret the evidence in order to develop these skills. Differences in performance on subtests should only be considered as meaningfully different if they are greater than or equal to six for the W-GCTA$^{UK}$ and five for the RANRA.

### 2.2.3 Outplacement and Career Guidance

The W-GCTA$^{UK}$ and RANRA can be used together in outplacement or career guidance. This might be appropriate for someone who faced redundancy, a change of circumstances, or was experiencing a lack of opportunity in a current role or profession and seeking an alternative. The purpose of the assessment process is to provide a wide perspective on suitable career paths and to help individuals choose options which best suit their abilities, needs and interests.

The W-GCTA$^{UK}$ and RANRA assess an aptitude for problem-solving, and allow an individual's potential to be explored, without pre-judging their suitability to a given role. This can help people develop an awareness of their own potential. This is useful for both those with and those without a clear idea about what to do.

Used together, the W-GCTA$^{UK}$ and RANRA produce two levels of information:

#### 2.2.3.1 Career scope and potential for analytical work

**Shaping up for Selection**

Jack had recently been made redundant and felt nervous about re-entering the job market; it had been 15 years since he had last undergone a selection process. At an outplacement counselling session he had a clear idea about the roles he wanted to go for.

The outplacement officer advised that he complete the RANRA and W-GCTA$^{UK}$ to prepare for assessment centres in the selection process.

Jack's scores were average (31st percentile for RANRA and the 49th percentile for W-GCTA$^{UK}$), but adequate, for the roles he was interested in.

Performance on the tests can indicate an individual's potential scope in a career. High scorers may be particularly suited to roles where there is a high need for analytical thinking and evaluation of data. These could include professional and high level strategic roles. Low scorers may be better suited to roles that do not rely heavily on these skills. These might include more operational duties where there is a much greater focus on interpersonal relationships.

Business Class

Corinne was studying Business at University and she attended the career's service to find out about graduate recruitment schemes in general management. The advisor reviewed her W-GCTA$^{UK}$ and RANRA scores and noted that she was above the 90th percentile on both tests in comparison to UK Management Trainees. He suggested that these scores were high enough to consider applying to high flier schemes run by graduate recruiters. Corinne was surprised, but very excited by this possibility and began to investigate employers running these schemes.

### 2.2.3.2 Type of analytical work to which the individual might be suited

In general, financial or scientific roles require higher levels of numerical reasoning than critical thinking. Those employed in roles that require problem solving and decision making using language and not numerical information need better critical thinking skills. Therefore, differences between scores on the W-GCTA$^{UK}$ and RANRA can help to identify the type of work individuals will be suited to. For example, if an individual achieves a significantly higher score on RANRA than on W-GCTA$^{UK}$, it can be inferred they may be more suited to a role that involves numerical reasoning, for example accounting, banking, analyst or financial roles. Alternatively, if an individual achieves a significantly higher score on the W-GCTA$^{UK}$ than RANRA, it can be inferred that they may be more suited to a role that involves critical thinking skills, for example marketing, legal advisor and human resources.

Care should be taken to avoid over-interpretation of test scores and differences between test scores. An individual's interests, motivations and circumstances will also be important factors in making career choices.

The W-GCTA$^{UK}$ and RANRA will not directly help an individual realise their ambition, but can provide critical information for evaluating possibilities available as well as an indication of potential success. Ultimately this should increase the chances of the individual fitting well into their next role.

Counselling Potential

Mal was made redundant after 30 years of employment in a factory. Mal had started his career as a Production Line Operator, but had quickly been promoted to the Production Line Manager. As part of the outplacement scheme Mal completed the W-GCTA$^{UK}$ and RANRA. The results showed that in comparison to the UK population, Mal's critical thinking skills were above average (75th percentile), while his numerical skills were average (40th percentile). Mal was surprised as he did not expect to do so well on the tests. After discussing these results with the Outplacement Consultant, Mal decided that he would be interested in re-entering education. With the facilitation of the Outplacement Consultant, Mal identified that he was interested in counselling. This was prompted by an interest in working with people and the skills he utilised in his recent role. Mal started volunteering with a local support group to gain some counselling experience and then applied for a full-time course, using his redundancy pay-out to fund his studies.

### 2.2.4 Training in Critical Thinking

Critical thinking is often taught in business and educational settings. The W-GCTA$^{UK}$ may be used to assess the extent to which trainees of these courses have mastered critical thinking skills. In this context the tests should be administered prior to and following completion of the training course. The time interval between the testing occasions should be carefully considered. The subscales may also provide useful diagnostic information on strengths and weaknesses.

If you require any assistance with this, please contact Pearson Assessment.

# 3. Choosing the W-GCTA$^{UK}$ and RANRA

In choosing to use the W-GCTA$^{UK}$ and RANRA, test users should be satisfied that these tests are relevant and appropriate to each situation they are to be used for. This will depend on the purpose of testing, the group being assessed and any practicalities, for example, time constraints. The section below provides guidance on choosing the W-GCTA$^{UK}$ and RANRA in different contexts and choosing an appropriate norm group with which to compare the test takers.

## 3.1 Choosing the W-GCTA$^{UK}$ and RANRA for Selection

In selection, the purpose of testing (and assessment) is to provide information needed to choose between job applicants. The information is collected for the employers use.

There are two key aspects to consider before using the W-GCTA$^{UK}$ or RANRA in selection:

1. Is an assessment of critical thinking skills and / or numerical reasoning relevant to the role?
2. If so, are the W-GCTA$^{UK}$ and/or RANRA relevant in terms of difficulty level and the group to be tested?

Job analysis provides recruiters with a clear understanding of a job and of what that job entails. Job analysis is the process of breaking down a job to its tasks, requirements, and performance criteria. There are formal methods of job analysis which are most effective (e.g. questionnaires, critical incident analysis) but as a minimum there should be a discussion with people who know the job well. It is advantageous to talk to both managers and job incumbents as they may have different perspectives on the role. Other informants may also be helpful (e.g. customers, trainers, reports). The information gathered is used to write a job description and person specification. A job description lists components of the job, duties or tasks, responsibilities and the required standards of performance. A person specification lays out the personal characteristics necessary to do the job; these include specific skills and abilities, interests, and disposition.

The information in the job description and person specification should be used to decide on the type of assessment that will be relevant to the role, for example critical thinking skills, numerical reasoning, spatial ability, mechanical aptitude, and will provide an indication of the level of difficulty of these characteristics.

The test user should confirm that the test is relevant in terms of its development, level and the group to be tested. This will include norms, reliability, validity, and group comparisons. Standards of assessment should not be higher than that which the job requires.

The test norms should provide suitable comparators for the intended test takers. These should contain an appropriate representative sample both with respect to the type of jobs applied for and background of the sample. Normed scores can be used to understand the ability level of a particular candidate or to rank order candidates for shortlisting purposes. However, it is not advised that shortlisting should be undertaken using a single measure as this will only reflect a single aspect of performance. Further information on using norms is provided at the end of this section.

Information on the reliability, validity and fairness of the W-GCTA$^{UK}$ and RANRA can be found in sections 5 and 6 of this guide. For further discussion, refer to the test manuals.

A good understanding of the role together with careful selection of tests and norm groups ensures sound evidence for the decision to use a test. Ideally this process should be documented. The organisation could be required to prove the relevance of any assessments if legally challenged.

### High Finance

A financial organisation wanted to identify employees for their High Fliers scheme. The programme was open to all current employees. Successful employees would be placed to allow them develop their skills and gain a broader perspective of the organisation. In addition, special training opportunities would be provided. The recruitment manager reviewed the job descriptions for all posts in which candidates would be placed. High level numerical reasoning was clearly required, as the roles involved analysis of financial markets, and the recruitment manager could see that RANRA would be appropriate. As this was a new scheme there was little information available about the details of the training courses. The recruitment manager spoke with trainers of relevant courses to get more information on the course requirements and discovered trainees are required to use and review technical written reports, for example legal and policy documents. Based on this evidence, the recruitment manager decided to use the W-GCTA$^{UK}$ as well as RANRA.

Technology Boost

Two years previously, a job analysis carried out for Business Analysts in a retail organisation identified key tasks to be financial reporting, forecasting sales and report writing. The organisation had been using RANRA and W-GCTA$^{UK}$ to assess for these roles.

Recently the organisation introduced some new technology which automatically reports and forecasts sales. The recruitment manager felt that numerical reasoning skills may be less important to this role than they once were. To gather more information, she spoke with the designers of the technology and two Sales Managers, who had been using the new technology for six months. She found that whilst Sales Managers no longer perform complex analysis of data they do need to be able to interpret the results. The recruitment manager decided that numerical reasoning was still necessary. She continued to use RANRA, but reduced the standard required to account for this change in the job and difficulty level.

### 3.2 Choosing the W-GCTA$^{UK}$ and RANRA for Development, Outplacement and Career Guidance

In guidance, the purpose of testing is to provide individuals with information they need to make realistic occupational decisions. Tests can be used to develop an awareness of skills and identify special training needs.

Tests should be used when they will help the individual to make decisions and explore broader options. The test must be used in agreement with the individual, and the test taker must be aware of the kind of information that can be gained, the decisions that are relevant, as well as the limitations of the test.

In choosing to use a test in this context the test user should evaluate what the test can do and the information being sought. The W-GCTA$^{UK}$ and RANRA allow individuals to develop an awareness of their potential; their ability to think critically and work with and evaluate numeric or verbal information. These tests allow a broad assessment for those who are unclear about what they want to do. For those who have a clear direction the tests provide a more specific assessment of their potential to succeed in relevant roles or training.

## 3.3 Choosing a Norm Group

A norm group provides a frequency distribution of scores from those who have previously taken the test. Using an appropriate norm group provides a meaningful standard for interpreting test scores. It allows an individual's score to be benchmarked against relevant others.

It is important that the comparison group used is appropriate for the test use. Where possible the comparison group should be taking the test for a similar purpose (e.g. selection, development). For career guidance it is often best to use the most general norm groups as these allow the person to benchmark their skills more effectively overall, whereas for selection the norm group should as far as possible reflect the selection context. In particular the comparison group should be applying for roles at a similar level although industry sector and job type are also important. Where the job level is not clear, typical educational background may provide an indication of this. Lastly, where possible the norms should reflect the diversity of the applicant sample with respect to gender, age, and ethnicity.

Organisations can use published or local norms:

### 3.3.1 Published Norms

Published norms allow an organisation to benchmark the performance of their employees or job applicants to that of others for a particular role (e.g. Senior Managers). The most recent norms are available on our website. Previously published norms are available in the W-GCTA$^{UK}$ and RANRA manuals.

### 3.3.2 Local Norms

Larger organisations can create a local or in-house norm group(s). These must be based on a sufficiently large group of people (at least 100) who are representative of the people being assessed. An in-house norm group will often be the most relevant comparison group. If there is insufficient data to create a local norm or the available group is unrepresentative in some way (e.g. scores come from graduate recruits and you wish to assess more experienced managers) it may be preferable to use a more general published norm group.

# 4. Using the W-GCTA$^{UK}$ and RANRA

Once you have decided to use the W-GCTA$^{UK}$ or RANRA you should follow the steps below.

## 4.1 Preparation for the Testing Session

- Ensure you are familiar with your organisation's Testing Policy.
- Schedule the testing sessions. Consider the duration of the session[1], accounting for preparation and instructions, and the number of people to be tested. Book an appropriate room as well as trained test administrators and additional invigilators (we recommend a ratio of at least 20:1 test takers to invigilators).
- Ensure the organisation has sufficient test materials in stock. You will need Test Booklets (one per candidate) and Record Forms (one per candidate). The Test Booklets may be re-used, however, you should hold spare copies of each in case of damage. The test log provides a checklist of materials needed (see appendix A). To order materials contact Pearson Assessment customer services.
- Invite test takers to the testing session. Inform them about the nature of tests, including how and why they are being used, the date, time, location and whether they are required to bring anything with them (e.g. some testing centres require personal identification to be checked).
- Distribute W-GCTA$^{UK}$ and RANRA familiarisation materials (if familiarisation materials are used, these should be provided to all test-takers).
- Find out about any disabilities and special requirements that any test takers have. Arrangements should be made to accommodate these. You should not change the standardised test administration procedure without taking advice from an expert as this can change the meaning of the scores. Contact Pearson Assessment for advice if you are unsure about making accommodations.
- Prepare the test log, this can act as a register and detail any reasonable adjustments to be made for candidates with disabilities as well as any unusual occurrences.

[1] Timed administration for the W-GCTA$^{UK}$ and the RANRA are 50 and 30 minutes respectively. Further details can be found in the W-GCTA$^{UK}$ and RANRA manuals.

## 4.2 Setting up the Testing Session

- Ensure all administrators and invigilators have appropriate training and are familiar with the W-GCTA$^{UK}$ and RANRA.
- Ensure a suitable room for testing, considering size, space, layout, lighting, temperature, noise and possible distractions. Test takers should be seated apart, but not directly opposite each other to avoid cheating and distraction. Ensure that potential disturbances are be minimised, e.g. phones are unplugged, 'Testing in Progress' signs are used.

## 4.3 Conducting the Testing Sessions

The testing session must be standardised to provide test takers with the same opportunity for doing well. It is advised that the testing instructions are closely followed. Try to create a friendly but purposeful atmosphere to put test takers at ease and enable them to work at their best. Start with an informal introduction to the testing session.

- An introduction should include:
  - Who you are
  - Your relationship to the organisation
  - The purpose of testing
  - How the results will be used
  - Who will have access to the results
  - Storage of the results (see Data Protection Act, 1998)[1]
  - What will happen after testing
  - The logistics of the testing session: breaks, fire alarms expected, duration, toilets
  - An opportunity for candidates to ask questions
- Ensure all mobile phones and electrical equipment are turned off and all candidates are ready to start the session
- On starting the session, ask test takers to maintain silence from this point on. They should raise their hand if they have any questions
- Follow standardised instructions (see appendices B, C and D)
- Ensure that the testing session is precisely timed. The start time should always be written on the test log
- Ensure that test takers are completing the Record Forms appropriately

[1] The Data Protection Act can be found at www.opsi.gov.uk.

- Ensure that all test materials are collected before anyone leaves the room
- Thank the test takers for attending and inform them of the next steps of assessment or process
- Complete test log
- Securely store test materials following testing session
- Ensure that the Data Protection Act is followed

## 4.4 Scoring the W-GCTA$^{UK}$ and RANRA

If you require the bureau service please contact Pearson Assessment. If you are scoring by hand, follow the three steps described below:

### Step 1 – Check Record Forms

1. Check each Record Form to ensure there are no multiple responses to the same item, missed items, or partly erased answers (where test takers have changed their response). Items that have been missed or have more than one marked response should not be scored. Where items are partly erased the test user must make a judgement as to whether to score the item and if so, which response to accept. Where it is clear that the test taker has made a considerable effort to erase or cross out one response and selected another, this item should be scored. Where it is unclear or both responses have been erased the item should not be scored.
2. Items that are not to be scored (i.e. those missed or with more than one response) should be crossed out with a line that will show through the acetate scoring key. This reduces error in hand scoring.

### Step 2 – Obtain the raw score

1. Place the acetate scoring keys over the Record Form
2. For each subtest count the correctly marked spaces
3. Record each subtest total in Box 1 of the 'Test Score Summary' table
4. Sum the subtest scores to create a total raw score
5. Record the total raw score in the 'Test Score Summary' table
6. Transfer the total raw score to Box A on the Record Form cover

*Note:*

The W-GCTA$^{UK}$ and RANRA award one point for every correct answer.

Do not count items as correct when more than one response has been selected, even though the right answer is one of those marked.

The maximum W-GCTA$^{UK}$ total raw score is 80, 16 for each subtest; the maximum RANRA total raw score is 32, 16 for each subtest.

### Step 3 – Convert raw scores into standardised scores using published norms

You will need a copy of the appropriate norm table. The latest norm tables can be found online, and previously published norms are available in the Test Manuals. Alternatively, you may have your own local norm group.

For the UK population norms (these can be found online or in the 2002 editions of the W-GCTA$^{UK}$ and RANRA manuals):

1. Look up each candidate's total raw score. This can be found in Box A on the Record Form cover.

2. Convert this score into a T score using Appendix A.

3. Use the T scores to look up percentile equivalents, stanines and stens using Appendix C.

For specific published norm groups (these can be found online or in the W-GCTA$^{UK}$ and RANRA manuals):

1. Convert the raw score to a percentile score (and/or standardised scores where presented) using the relevant norm table

## 4.5 Interpreting scores

Test scores should be interpreted in the given context, against appropriate norm groups and related to additional information. It is important to relate the scores to the purpose of testing, making appropriate connections between the test score and what this actually means in terms of a career or a specific job.

Test scores provide an indication of an individual's performance against that of a group of others who have taken the test. To allow ease in the comparison of this individual against others, test scores are standardised. A description of standardised scores is presented in the following pages.

### 4.5.1 Standardised Scores

#### T scores

T scores provide a standardised scale for comparing scores. T scores have a mean of 50 and a standard deviation of 10 which means that with a normal distribution of scores, 68% of test takers will score between 40 and 60. The advantage of T scores is that they represent an even scale. That is, the difference between scores of 70 and 80 is the same as the difference between scores of 45 and 55. In addition, it is possible to apply a standard error of measurement to a T score to allow for calculation of a band of error around a score. Generally, T scores should not be used in feedback unless the test taker has an understanding of psychometrics and statistics.

#### Percentiles

Percentile equivalent scores are often used when giving feedback to test takers. These have the advantage of being readily understood and allow test takers to appreciate how well they have done in comparison to others. It is important not to confuse percentiles with percentages. A percentile expresses a score by the percentage of test takers in the norm group who scored less. This means that a test taker who scores at the 70th percentile has scored higher than 70 percent of the comparison group. A score at the 30th percentile is better than the 30 percent of the comparison group.

Percentiles are not equal units. They show the relative position or ranking of each test taker in comparison to the norm group, but do not illustrate the amount of difference between scores. In a normal distribution, cases will be clustered more closely at the centre of the distribution than at the extremes. Differences at this mid-point are therefore more exaggerated while those at the extremes are relatively understated.

#### Banded or graded scores:

To simplify scores further they may be banded or graded into the following categories:

A / 1 – Well above average performance, 91st percentile and above

B / 2 – Above average performance, 71st – 90th percentiles

C / 3 – Average, 31st – 70th percentiles

D / 4 – Below average performance, 11th – 30th percentiles

E / 5 – Well below average performance, 10th percentile and below

### Stens and Stanines

Stens are standardised scores which have a 10-point scale ranging from 1 to 10. Stens have a mean of 5.5 and a standard deviation of 2. Stanine scores are a slight variation on stens, ranging from 1 to 9, with a mean of 5 and a standard deviation of 2. Both stens and stanines are commonly used in feedback.

### 4.5.2 Accuracy of test scores

Scores obtained on the W-GCTA$^{UK}$ and RANRA, or any other psychological test, can only be considered an estimate of the test takers true score. No test is perfectly accurate (without error), and the standard error of measurement (SEM) indicates the amount of error to be expected in a test takers score. This means that scores should be interpreted as bands, not precise points.

Banding of scores serves to check against over-emphasising small differences between scores. The standard error of measurement (SEM) can be used to create a band for a score when expressed as a raw score or a T score. A score should be considered to fall within a band from one SEM below the score to one SEM above the score 68 percent of the time.

> Banding Scores
>
> Femi has a T score of 45, John has T score of 53 and Juliette has a T score of 55 on the RANRA. The T score SEM for RANRA is 4.5. Therefore Femi's score can be considered to lie between 40.5 and 49.5. John's ranges between 48.5 and 57.5 and Juliette's should be considered to be between 50.5 and 59.5.

### 4.5.3 Comparing scores between test takers on the same test

When comparing the performance of two individuals on one test the standard error of difference (SE(diff)) is used to judge whether the scores are significantly different from one another. The SE(diff) is equivalent to 1.4 times the SEM when comparing two scores on the same test. The SEM for the W-GCTA$^{UK}$ is 4.8 in T scores so the SE(diff) is 6.7. Two SE(diffs) are required for 95 percent confidence.

### Score Differences

An organisation recruiting a senior manager held an assessment centre which consisted of the W-GCTA$^{UK}$, a personality questionnaire and feedback interview, and a case study exercise. Two candidates obtained T scores of 75 and 71 on the W-GCTA$^{UK}$. The SEdiff for the W-GCTA$^{UK}$ is 6.7. The recruitment manager decided that these scores were not significantly different. The recruitment manager was confident that the applicants did not differ in terms of their critical thinking skills but was reassured to know that both applicants met the minimum level.

#### 4.5.4 Comparing W-GCTA$^{UK}$ and RANRA scores for a test taker

When comparing scores from W-GCTA$^{UK}$ and RANRA for the same individual the standard error of difference (SEdiff) is important. For T scores for the two tests the standard error of difference is 6.4. To be certain (95% confident) that the two scores are different, you should allow two standard errors of difference (about 13 T score points) between scores.

### Test Differences

Tim has a W-GCTA$^{UK}$ T score of 60 and a RANRA T score of 69. These scores do not differ significantly. Leigh has a W-GCTA$^{UK}$ T score of 60 and a RANRA T score of 78. These scores are significantly different, greater than 13 points apart, so it can then be assumed that Leigh's numerical reasoning skills are significantly stronger than his

## 4.6 Limitations of test scores

Test scores should be interpreted carefully. Errors may arise in the administration of the testing session or in scoring. Scores can also be affected by a test taker's state, for example, anxiety or feeling unwell. Candidates with a disability or with English as a second language may be disadvantaged due to the test format. For these reasons scores should be explored carefully and interpreted with caution.

On occasion test scores may contradict alternative information on a test taker. In this case, the test user should work with the test taker to explore the information and discover possible causes for these anomalies.

The W-GCTA$^{UK}$ and the RANRA have been carefully standardised using standard administration procedures. Any changes to this process can reduce the reliability of the test scores.

Used correctly psychometric tests are a powerful tool that can provide important information on the test taker but because of these limitations tests are designed to be used alongside additional assessment methods.

## 4.7 Feedback of W-GCTA$^{UK}$ and RANRA Test Scores

It is best practice to provide test takers with appropriate feedback on their performance in a psychometric test or assessment process. This also serves to increase the perceived acceptability of the test.

Feedback is an essential process by which to inform test takers of their test performance and its implications, in a language they can understand. Information given should be fair, accurate, understandable and any questions should be answered. Providing feedback to the test taker can be a sensitive process as some people have emotional reactions to information about their strengths and weaknesses.

Feedback can be written, face-to-face or over the telephone. Written feedback may be appropriate where there are a large number of test takers and face-to-face or telephone feedback will not be feasible or cost effective.

Pearson Assessment are able to provide computer generated reports for both the W-GCTA$^{UK}$ and RANRA. Examples of these can be found in appendix E.

In preparation for feedback, the qualified person providing feedback should:

- Consult the test log to establish if there were any problems or interruptions that may have affected test performance. Interpretation of test scores assumes standardisation of test conditions, and so a fair and accurate assessment will rely on this
- Ensure that scores are translated into the appropriate standard scores or percentiles using a relevant norm group
- Check that the test taker has a clear understanding of the relevance of the W-GCTA$^{UK}$ and RANRA in the context it has been used

## Using appropriate scores in feedback

The approach to feedback that is adopted should be the one that produces the most appropriate outcome from the test taker's point of view.

Scores that are accessible to the test taker should be used; the most commonly used scores for feedback include percentiles, stens and stanines.

Feedback should allow the test taker to understand the purpose and relevance of the assessment, as well as how they performed in the test. This should follow three steps:

1. Describe the tests used and the purpose of the assessment,

   For example:

   "You recently completed the W-GCTA$^{UK}$ and RANRA as part of a recruitment process. The W-GCTA$^{UK}$ consists of five components or subtests that are designed to assess critical thinking, that is, your ability to critically evaluate written information. The RANRA consists of two subtests that assess numerical reasoning, your ability to analyse and evaluate numeric information. The scores were used together with other information in making selection decisions".

2. Describe the individual's results in the context of the comparison group. Explain what the group is, why it was selected, why it is relevant and how the individual's performance compared to the group.

   For example:

   "Compared with the broad based representative group from the working population who have completed this test, your score was at the 80th percentile. That is, you did better than 80 percent of others who have taken the test. Only 20 percent of test takers did better than you."

   Or,

   "Compared with others who have applied for the position of underwriter in this organisation, your score was at the 45th percentile. That is, you did better than 45 percent of others who have taken the test for this purpose."

   Or

   "Your score fell in Band 3 (or Grade C). This means that your performance can be classified as average in the group of police officers with which you were compared."

3. Describe the relevance of the scores for the purpose in which they are being used.

   For example:

   "Critical thinking skills are crucial to the role of legal advisor. The role holder will be required to use these skills to analyse written information from time to time in the job."

Good feedback should:

- Put the test taker at ease
- Be pitched at the appropriate level for the test takers knowledge of psychometric assessment, the W-GCTA$^{UK}$ and/or RANRA
- Provide relevant information about the test
- Describe the group against which the test taker is being compared
- Describe performance in relation to that group
- Avoid technical terms or jargon
- Place test results in the context of other information gained during assessment
- Provide test takers with the opportunity to ask questions
- Be a positive experience for test takers, where information is related to their needs

Feedback should always be meaningful to the test taker. In selection, this is reporting how the applicant performed in comparison to the norm group used. In development or guidance, it may be useful to compare the test takers score against a range of norms to ensure that it is placed in context and fully understood.

For example, a candidate with a W-GCTA$^{UK}$ raw score of 53 is classified as well-below average in comparison with UK Banking applicants (5th percentile), but average compared to the UK general population (34th percentile). Without this additional information the test taker may be misinformed about their level of critical thinking.

# 5. Reliability and Validity

## 5.1 Reliability

A reliable test provides an accurate measurement that is consistent and stable over time. A test administrator's contribution to reliability is crucial. To maintain reliability a test should be administered following the publisher's instructions closely and carefully.

Two measures of reliability are discussed here: internal consistency and test-retest reliability.

### 5.1.1 Internal Consistency

The coefficient of internal consistency reflects the accuracy of scores on a test. The W-GCTA$^{UK}$ and RANRA both have good internal consistency, 0.81 and 0.78 respectively, for the standardisation sample of 1546 individuals (see W-GCTA$^{UK}$ and RANRA manuals for details regarding the composition of this sample).

### 5.1.2 Test-retest Reliability

The test-retest reliability coefficient for the W-GCTA$^{UK}$ was 0.73 for a group of 96 students, with a 3 month interval between testing. Values over 0.70 reflect an acceptable measure of stability over time.

## 5.2 Validity

A valid test effectively measures what it intends to measure and 'gives the information a decision maker needs' (Cronbach, 1970).

Three forms of validity are examined here: content validity, construct validity and criterion validity.

### 5.2.1 Content Validity

Content validity relates to the degree that the content of the test reflects what is to be measured. Evidence for content validity is established through test construction and expert judgements on the relevance of test items and the nature of the tasks.

The W-GCTA$^{UK}$ offers a valid estimate of proficiency in critical thinking (Houle, 1943; Morse & McCune, 1957; Paul and Nosich, 1992), and has been judged to represent an adequate sample of critical thinking skills; for example, those laid out by the Cooperative Study of Evaluation in General Education (1954):

1. the ability to define a problem

2. the ability to select pertinent information for the solution of the problem

3. the ability to recognise stated and unstated assumptions

4. the ability to formulate and select relevant and promising hypotheses

5. the ability to draw conclusions validly and to judge the validity of inferences

The RANRA was constructed emphasising estimation and analysis in numerical reasoning. The subtests require deduction, interpretation and evaluation in problem solving with numerical information; these skills are comparable with 2, 4 and 5 in the above description of critical thinking (Cooperative Study of Evaluation in General Education, 1954). In the workplace, calculations are now carried out by technology and people are required to consider the suitability of these responses. RANRA can be judged to assess this.

In the employment setting, evidence of content validity is demonstrated by measuring competencies that are relevant to the job. Critical thinking and numerical reasoning, measured by the W-GCTA$^{UK}$ and RANRA, are relevant to a wide range of jobs. The relevance of the W-GCTA$^{UK}$ and RANRA to a particular job can be confirmed by a clear understanding of the knowledge, skills and abilities required in that job; guidance on this is provided in section 3.1 of this guide.

### 5.2.2 Construct Validity

Construct validity is the extent to which the measure can be shown to respond as it should, given what it is intended to measure. Evidence of construct validity can be established through examination of the relationship between a test and other assessments.

The W-GCTA has significant relationships with a number of tests of reasoning ability and assessment centre exercises measuring cognitive skills, ranging from 0.26 - 0.81. Details of these studies are presented in appendix F.

### 5.2.3 Criterion Validity

Criterion validity is established when performance on a test is related to performance on an external criterion, for example job performance or training success.

The W-GCTA$^{UK}$ and RANRA have been used to predict performance according to a variety of criteria in employment and training settings. A summary of some of these studies can be found in appendix G.

#### 5.2.3.1 Job Performance

The W-GCTA has been found to predict success across a range of occupations, including analysts (Ejiogu, Yang, Trent & Rose, 2006), management trainees (Bell, 1998) and pilots (Mosher 1999).

More specifically the W-GCTA$^{UK}$ has been found to relate to level of management; success, based on track record, in senior management roles (Crump & Knight, 1999); and performance in executive management duties (Gaston, 1993).

Performance on the W-GCTA$^{UK}$ has also been linked to salary (Hildebrandt & Lucas, 1980).

The W-GCTA$^{UK}$ has been used far more widely and frequently than the RANRA, and hence more research is available. One study conducted by Ejiogu, Rose, Yang & Trent (2007) showed that the W-GCTA$^{UK}$ and RANRA were both predictors of job performance and job potential in finance-related positions, and that RANRA significantly contributed to this process.

#### 5.2.3.2 Training Outcomes

The W-GCTA$^{UK}$ has been found to predict training outcomes in a number of settings, for example, medical courses (Behrens 1996; Miller, Sadler & Mohl, 1993; Molidor, Elstein & King 1978), professional training courses (Gadzella, Stacks, Stephens & Masten, 2005), general academic success (McCammon, Golden & Wuensch, 1988; Steward & Al Abdulla, 1989), teaching critical thinking skills (Heraty & Morley, 2000; Wood, 1981); and educational programmes (Williams, 2003; Sandor, Clark, Campbell, Rains & Cascio, 1998; Hurov, 1987; Wilson & Wagner, 1981; Hildebrandt & Lucas, 1980).

### 5.2.4 Validity Generalisation

The principle of validity generalisation refers to the extent that inferences from previous validity research can be generalised to other instruments and situations.

There is a consistent body of research that suggests that cognitive ability tests, like the W-GCTA$^{UK}$ and RANRA, are the single most effective predictor of job performance (e.g. McDaniel, *et al*., 1994; Ones *et al.*, 1994).

### 5.2.5 Local validation

It is important to conduct local validation where possible. Organisations seeking assistance in local validation should contact Pearson Assessment.

## 6. Group Comparisons

Research shows that differences in test performance between groups can occur. These differences vary by the test and the group of interest. Where test score differences reflect true differences in an ability which is relevant to job performance, a test should be considered as fair. Where test scores differences do not reflect differences in ability or cannot be linked to job performance, a test may be considered unfair.

The W-GCTA$^{UK}$ did not show any gender differences in the UK standardisation. Findings on age and the W-GCTA$^{UK}$ are in line with the research on age and critical thinking skills.

The average performance of men typically exceeds that of women in numerical reasoning (Casey *et al.*, 1995; Fivush & Golombok, 1994). Numerical scores can decline where candidates are very out of practice in working with numbers. Therefore, users are strongly advised always to provide suitable practice materials. Findings on gender and age differences in RANRA scores, from the UK standardisation, are small but significant and reflect previous research.

Differences between some ethnic groups are typically found for the W-GCTA$^{UK}$ or RANRA similar to those found in other reasoning tests, but are at the higher end. Therefore, it is important for users to only use tests when they are clearly relevant, and to be cautious in setting cut scores, as adverse impact is reduced with lower test cut scores.

# Appendix A

A test log should always be maintained. An example is provided below. This page may be photocopied for use with the W-GCTA$^{UK}$ and RANRA in your organisation.

## Test Log

**Organisation:** _____

**Purpose of Testing:** Selection / Development / Appraisal

**Test(s) Used:** W-GCTA$^{UK}$ / RANRA

**Test Administrator:** _____

**Test Invigilators:** _____

**Date:** _____

**Start Time:** _____

**Finish Time:** _____

## Candidate List

| | |
|---|---|
| 1. | 6. |
| 2. | 7. |
| 3. | 8. |
| 4. | 9. |
| 5. | 10. |

PEARSON

## Materials Checklist

| | Number checked out | Number checked in |
|---|---|---|
| W-GCTA$^{UK}$ Test Booklets: | | |
| W-GCTA$^{UK}$ Record Forms: | | |
| RANRA Test Booklets: | | |
| RANRA Record Forms: | | |
| Test Instructions: | | |
| Stopwatches: | | |
| Pencils: | | |
| Erasers: | | |
| Pencil Sharpeners: | | |

## Disturbances / Unusual Occurrences

PEARSON

# Appendix B

## W-GCTA$^{UK}$ Test Administration Instructions

Write your notes on an informal introduction to the testing session here:

Following an introduction to the testing session, say:

From now on, please do not talk among yourselves, but ask if anything is not clear.

Distribute the Test Booklets and say:

Do not open these booklets until you are told to do so.

Then distribute the Record Forms and say:

Please complete the candidate details on the cover of this form.

If you are collecting equal opportunities data say:

Equal opportunities data are collected to monitor fairness in testing. Completion of this section is optional.

Otherwise say:

You do not need to complete the equal opportunitites section.

Allow the test takers time to complete the details on the cover of the Record Form. Then say:

In this test, all the questions are in the Test Booklet. There are five separate parts to the test in the booklet and each one is preceded by its own directions and examples. For each question, decide what you think is the best answer. As your score will be derived from the number of items you answer correctly, try to answer each question even if you are not sure if the answer is correct. Record your

choice by putting a cross in the appropriate place on the Record Form. Always be sure that the answer space has the same number as the question in the Test Booklet. Do not make any other marks on the Record Form. If you change your mind about an answer make sure that you rub out the first mark completely. Do not spend too much time on any one question. When you finish a page, go straight on to the next one, working through each of the tests in turn. If you finish all of the tests before the time is up, you may go back and check your answers.

Say:

You will have 50 minutes[1] to work on the test. Now read the directions on the cover of your Test Booklet.

After allowing time for test takers to read the directions, say:

Are there any questions about what you are to do?

Answer any questions, preferably by re-reading the appropriate sections of the directions, then say:

Ready? ... Begin.

Immediately start your timing procedure. If any of the examinees finish before the end of the test period, either tell them to sit quietly until everybody has finished or collect their materials and dismiss them quietly.

While the group is taking the test, move about the room making sure that each test taker is marking the Record Form properly.

At the end of 50 minutes say:

Stop! Put your pencils down. This is the end of the test.

### Concluding administration

At the end of the testing session, collect the Test Booklets, Record Forms and pencils and thank everyone for attending.

The W-GCTA$^{UK}$ is a demanding test to take. The style of the items in some of the subtests makes it difficult for test takers to achieve a confident appreciation of their performance in the test. From this point of view it can be an uncomfortable experience and some words of reassurance at this point may be appropriate. It may be constructive to clarify the contribution of the test within the context of other aspects of selection or appraisal procedures. It would also be constructive to reassure test takers regarding the confidentiality of test scores.

[1] There are circumstances where this time may be varied or an untimed testing session is administered (see section 3.4 of the W-GCTA$^{UK}$ Manual).

# Appendix C

## RANRA Test Administration Instructions

Write your notes on an informal introduction to the testing session here:

Following an introduction to the testing session, say:

From now on, please do not talk among yourselves, but ask if anything is not clear.

Distribute the Test Booklets and say:

Do not open these booklets until you are told to do so.

Then distribute the Record Forms and say:

Please complete the candidate details on the cover of this form.

If you are collecting equal opportunities data say:

Equal opportunities data are collected to monitor fairness in testing. Completion of this section is optional.

Otherwise say:

You do not need to complete the equal opportunitites section.

Allow the test takers time to complete the details on the cover of the Record Form. Then say:

In this test, all the questions are in the Test Booklet. There are two separate tests in the booklet and each one is preceded by its own directions and examples. For each question, decide what you think is

the best answer. As your score will be derived from the number of items you answer correctly, try to answer each question even if you are not sure if the answer is correct. Record your choice by putting a cross in the appropriate place on the Record Form. Always be sure that the answer space has the same number as the question in the Test Booklet. Do not make any other marks on the Record Form. If you change your mind about an answer make sure that you rub out the first mark completely. Do not spend too much time on any one question. When you finish a page, go straight on to the next one, working through each of the tests in turn. If you finish all of the tests before the time is up, you may go back and check your answers.

Say:

You will have 30 minutes[2] to work on the test. Now read the directions on the cover of your Test Booklet.

After allowing time for test takers to read the directions, say:

Are there any questions about what you are to do?

Answer any questions, preferably by re-reading the appropriate sections of the directions, then say:

Ready? ... Begin.

Immediately start your timing procedure. If any of the examinees finish before the end of the test period, either tell them to sit quietly until everybody has finished or collect their materials and dismiss them quietly.

While the group is taking the test, move about the room making sure that each test taker is marking the Record Form properly.

At the end of 30 minutes say:

Stop! Put your pencils down. This is the end of the test.

### Concluding administration

At the end of the testing session, collect the Test Booklets, Record Forms and pencils and thank everyone for attending.

The RANRA is a demanding test to take. The style of the items in some of the subtests makes it difficult for test takers to achieve a confident appreciation of their performance in the test. From this point of view it can be an uncomfortable experience and some words of reassurance at this point may be appropriate. It may be constructive to clarify the contribution of the test within the context of other aspects of selection or appraisal procedures. It would also be constructive to reassure test takers regarding the confidentiality of test scores.

[2] There are circumstances where this time may be varied or an untimed testing session is administered (see page 3.4 of the RANRA Manual).

# Appendix D

## Administering the W-GCTA$^{UK}$ and RANRA within a single session

### W-GCTA$^{UK}$ and RANRA Test Administration Instructions

Write your notes on an informal introduction to the session here:

Following an introduction to the testing session, say:

From now on, please do not talk among yourselves, but ask if anything is not clear.

Distribute the Test Booklets and say:

Do not open these booklets until you are told to do so.

Then distribute the Record Forms and say:

Please complete the candidate details on the cover of the W-GCTA$^{UK}$ Record Form.

If you are collecting equal opportunities data say:

Equal opportunities data are collected to monitor fairness in testing. Completion of this section is optional.

Otherwise say:

You do not need to complete the equal opportunitites section.

Allow the test takers time to complete the details on the cover of W-GCTA$^{UK}$ Record Form. Then say:

In this test session, all the questions are in the two Test Booklets. There are five separate tests in the first booklet, W-GCTA$^{UK}$, and two separate tests in the second booklet, RANRA. Each of the seven tests is preceded by its own directions and examples.

PEARSON

Each test booklet has a time limit. You will be given 50 minutes to work through the first booklet, W-GCTA$^{UK}$. I will then ask you to stop and collect the booklet together with your completed Record Form.

The test session will resume with the second booklet, RANRA, which you will have 30 minutes to complete.For each question, decide what you think is the best answer. As your score will be derived from the number of items you answer correctly, try to answer each question even if you are not sure if the answer is correct. Record your choice by putting a cross in the appropriate place on the Record Form. Always be sure that the answer space has the same number as the question in the Test Booklet. Do not make any other marks on the Record Form. If you change your mind about an answer make sure that you rub out the first mark completely. Do not spend too much time on any one question. When you finish a page, go straight on to the next one, working through each of the tests in turn. If you finish all of the tests before the time is up, you may go back and check your answers.

Please sit patiently until the end of the timed session and do not disturb others around you.

Say:

You will have 50 minutes[3] to work on the first test, W-GCTA$^{UK}$ containing five subtests.

Now read the instructions on the cover of your Test Booklet.

After allowing time for test takers to read the directions, say:

Are there any questions about what you are to do?

Answer any questions, preferably by re-reading the appropriate sections of the directions, then say:

Ready? ... Begin.

Immediately start your timing procedure. If any of the examinees finish before the end of the 50 minute test period, tell them to sit quietly until everybody has finished.

While the group is taking the test, move about the room making sure that each test taker is marking the Record Form properly.

At the end of 50 minutes say:

Stop! Put your pencils down. This is the end of the first five tests

Collect the W-GCTA$^{UK}$ Test Booklets and Record Forms

Say:

Please complete the candidate details on the cover of the RANRA Record Form.

[3] There are circumstances where this time may be varied or an untimed testing session is administered (see section 3.4 of the W-GCTA$^{UK}$ and RANRA Manuals).

If you are collecting equal opportunities data on the RANRA say:

Again, completion of the equal opportunities section is optional.

Otherwise say:

You do not need to complete the equal opportunities section[4].

Allow the test takers time to complete the details on the cover of W-GCTA$^{UK}$ Record Form. Then say:

For each question, decide what you think is the best answer. As your score will be derived from the number of items you answer correctly, try to answer each question even if you are not sure if the answer is correct. Record your choice by putting a cross in the appropriate place on the Record Form. Always be sure that the answer space has the same number as the question in the Test Booklet. Do not make any other marks on the Record Form. If you change your mind about an answer make sure that you rub out the first mark completely. Do not spend too much time on any one question. When you finish a page, go straight on to the next one, working through each of the tests in turn. If you finish all of the tests before the time is up, you may go back and check your answers.

Say:

You will have 30 minutes[5] to work on the second test, RANRA, containing two subtests.

Now read the instructions on the cover of your Test Booklet.

After allowing time for test takers to read the directions, say:

Are there any questions about what you are to do?

Answer any questions, preferably by re-reading the appropriate sections of the directions, then say:

Ready? ... Begin.

Immediately start your timing procedure. If any of the examinees finish before the end of the tests period, either tell them to sit quietly until everybody has finished or dismiss them quietly.

While the group is taking the test, move about the room making sure that each test taker is marking the Record Form properly.

At the end of 30 minutes say:

Stop! Put your pencils down. This is the end of the test.

[4] If you have collected equal opportunities data on the W-GCTA$^{UK}$ it may be unnecessary for test takers to completet the RANRA equal opportunities section. Please note, the equal opportunities section may differ between reprints as we are constantly updating our forms.

[5] There are circumstances where this time may be varied or an untimed testing session is administered (see section 3.4 of the W-GCTA$^{UK}$ and RANRA Manuals).

## Concluding administration

At the end of the testing session, collect the Test Booklets, Record Forms and pencils and thank everyone for attending.

The W-GCTA$^{UK}$ and RANRA are demanding tests to take. The style of the items in some of the subtests makes it difficult for test takers to achieve a confident appreciation of their performance in the test. From this point of view it can be an uncomfortable experience and some words of reassurance at this point may be appropriate. It may be constructive to clarify the contribution of the test within the context of other aspects of selection or appraisal procedures. It would also be constructive to reassure test takers regarding the confidentiality of test scores.

# Appendix E

## Example Feedback Reports

| | |
|---|---|
| W-GCTA$^{UK}$ | Low Score |
| | High Score |
| RANRA | Average Score |
| | High Score |

The following reports show scores Corrected for Guessing. Please see W-GCTA$^{UK}$ Manual 7.15 and RANRA Manual 7.11.

# Appendix F

## Summary of Construct Validity Studies

| Study | Test | | Sample | N | Comparison Test | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | W-GCTA | RANRA | | | Verbal | Numerical | Non-verbal | Achievement Test | IQ |
| Moutafi, Furnham & Crump, 2003 | ✓ | | UK | 900 | | | 0.39 | | |
| 1994 W-GCTA manual | ✓ | | US | | 0.66, 0.51 | 0.41 | 0.26 | | |
| 1994 W-GCTA manual | ✓ | | US | | | | 0.40 | | |
| McMurray, Beisenherz & Thompson (1991) | ✓ | | US | 23-47 | | | 0.40 | 0.64; 0.85; 0.78 | |
| 1980 W-GCTA Manual | ✓ | | US | 31-55 | | | | 0.50; 0.30; 0.49; 0.24; 0.51; 0.53; 0.32; 0.35 | 0.62; 0.41; 0.70; 0.61; 0.70; 0.78; 0.62; 0.81; 0.70 |
| 1980 W-GCTA Manual | ✓ | | US | 20-23 | | | | 0.69; 0.37 | |
| 1980 W-GCTA Manual | ✓ | | US | 61 | | | | 0.65; 0.64 | |
| 1980 W-GCTA Manual | ✓ | | US | 5,237 | | | | | 0.60 |
| 1980 W-GCTA Manual | ✓ | | US | 378 | | | | 0.54; 0.43 | |
| 1980 W-GCTA Manual | ✓ | | US | 60 | | | | 0.60; 0.41 | |
| 1980 W-GCTA Manual | ✓ | | US | 24-32 | | | | 0.45; 0.48; 0.21; 0.30 | |
| 1964 W-GCTA Manual | ✓ | | US | 167 | 0.70 | | 0.43 | | |
| 1964 W-GCTA Manual | ✓ | | US | 20,312 | | | | | 0.75; 0.73 |

# Appendix G

## Summary of Criterion Validity Studies

| Study | Test | | Sample | N | Job Performance | | | Assessment Centre Ratings | | Educational Achievement |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Management ratings | | Other | | | |
| | W-GCTA | RANRA | | | Cognitive | Overall | | Cognitive | Non-Cognitive | |
| Ejiogu, Rose, Yang & Trent (2007) | | ✓ | US | 87 | | 0.37, 0.37, 0.33 | | | | |
| Ejiogu, Rose, Yang & Trent (2007) | ✓ | | US | 87 | | 0.34, 0.29, 0.18 | | | | |
| Ejiogu, Yang, Trent & Rose (2006) | ✓ | | US | 59-66 | 0.52; 0.52; 0.48 | 0.51; 0.32 | | | | |
| Gadzella, Hogan, Masten, stacks, Stephens and Zascavage (2006) | ✓ | | US | 56-565 | | | | | | 0.30, 0.62, 0.38, 0.24, 0.20, 0.36 |
| Gadzella, Stacks, Stephens & Masten (2005) | ✓ | | US | | | | | | | 0.31 |
| Williams (2003) | ✓ | | US | 428 | | | | | | 0.42, 0.57 |
| Kudish & Hoffman (2002) | ✓ | | US | 71 | | | | 0.58; 0.43 | | |
| Spector, Schneider, Vance & Hezlett (2000) | ✓ | | US | 429 | | | | 0.26 | 0.16 | |
| Behrens 1996 | ✓ | | US | | | | | | | 0.59; 0.53; 0.51 |

| Study | Test | | Sample | N | Job Performance | | | Assessment Centre Ratings | | Educational Achievement |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Management ratings | | Other | | | |
| | W-GCTA | RANRA | | | Cognitive | Overall | | Cognitive | Non-Cognitive | |
| 1994 W-GCTA manual | ✓ | | US | 141-241 | | | 0.25; -0.05 | | | |
| 1994 W-GCTA manual | ✓ | | US | 201 | | | -0.06; -0.19 | | | |
| 1994 W-GCTA manual | ✓ | | US | 425-430 | | | 0.18; 0.01 | | | |
| Bauwens & Gerhard (1987) | ✓ | | US | 159 | | | | | | 0.30; 0.32; 0.42; 0.26; 0.24; 0.50 |
| Gross, Takazawa & Rose (1987) | ✓ | | US | 45 | | | | | | 0.32; 0.35; 0.38 |
| Holgren & Covin (1984) | ✓ | | US | 60 | | | | | | 0.50; 0.46 |
| 1980 W-GCTA manual | ✓ | | US | 26-34 | | | | | | 0.16; 0.35 |
| 1980 W-GCTA manual | ✓ | | US | 61 | | | | | | 0.30 |
| 1980 W-GCTA manual | ✓ | | US | 116 | | | | | | 0.12 |
| 1980 W-GCTA manual | ✓ | | US | 86 | | | | | | 0.19 |
| Kooker (1971) | ✓ | | US | 69 | | | | | | 0.37 |

## References


Behrens, P.J. (1996). The Watson-Glaser Critical Thinking Appraisal and academic performance. Journal of Nursing Education, 35, 34-36.

Bauwens, E.E. & Gerherd, G.G. (1987). The use of the Watson-Glaser Critical thinking Appraisal to predict success in a baccalaureate-nursing program. Journal of Nursing Education, 26, 278-281.

Bell, W.A. (1998). An assessment of the relationship between critical thinking success in presolo flight training. Dissertation Abstract international, 59 (5-A), 1803.

Casey, M.B., Nutall, R., Pezaris, E., Persson Benbow, C. (1995). The influence of spatial ability on gender differences in mathematics college entrance test scores across diverse samples. Developmental Psychology, 31, 697-705.

Dressel, P., & Mayhew, L. (1954). General Education: Exploration in Evaluation. Final Report of the Cooperative Study of Evaluation in General Education. Washington, D.C: American Council on Education.

Cronbach, L.J. (1970). Essentials in psychological testing (3rd ed). New York: Harper & Row.

Crump, J., & Knight, M. (1999). Senior managers: are they really different? Retrieved 16 June 2008, from http://www.kaisen.co.uk/pdf/senior-managers-different-new.pdf

Ejiogu, K.C., Rose, M., Yang, Z., & Trent, J. (2007). Incremental Validity of Numerical Reasoning over Critical Thinking. Retrieved 16 June 2008, from http://harcourtassessment.com/NR/rdonlyres/FE53E240-6C90-4064-B90A-5F20E789AA05/0/NumericReason.pdf

Ejiogu, K.C., Yang, Z., Trent, J., & Rose, M. (2006). Understanding the relationship between critical thinking and job performance. Retrieved 16 June 2008, from http://harcourtassessment.com/NR/rdonlyres/6E133ED3-5516-463A-B880-D06C235A06E0/0/CritcalThinking.pdf

Gadzella, B.M., Stacks, J., Stephens, R.C., & Masten, W.G. (2005). Watson-Glaser Critical Thinking Appraisal, Form S for education majors. Journal of Instructional Psychology, 32, 9-12.

Gadzella, B.M., Hogan, L., Masten, W., Stacks, J., Stephens, R., & Zascavage, V. (2006). Reliability and validity of The Watson-Glaser Critical Thinking Appraisal-Forms for different academic groups. Journal of Instructional Psychology, 33 (2), 141-145.


PEARSON

Gaston, A. (1993). Recognizing potential law enforcement executives. (Reports No. NCJ 131646) Washington, D.C.: National Institute of Justice/NCJRS.

Golombok, S., & Fivush, R. (1994). Gender development. New York: Cambridge University Press.Golombok, S., & Fivush, R. (1994). Gender development. New York: Cambridge University Press.

Gross, Y.T., Takazawa, E.S., & Rose, C.L. (1987). Critical thinking in nursing education. Journal of Nursing Education, 26, 317-323.

Heraty, N., & Morley, M.J. (2000). The application of the structure of intellect programme: A manufacturing facility experiment. Journal of Managerial Psychology, 15, 691-711.

Hildebrandt, S., & Lucas, J.A. (1980). Follow-up of students who majored and are majoring in legal technology. Research report, William Raney Harper College, Pallatine, IL.

Holgren, B., & Covin, T. (1984). Selective characteristics of preservice professionals. Education, 104, 321-328.

Houle, C. (1943). Evaluation in the eight-year study. Curriculum Journal, 14, 18-21.

Hurov, J.T. (1987). A study of the relationship between reading, computational, arid critical thinking skills and academic success in fundamentals of chemistry. (ERIC Document Reproduction Service No. ED 286569).

Kooker, E.W. (1971). The relationship between performance in a graduate course in statistics and the Miller Analogies Tests and the Watson-Glaser Critical Thinking Appraisal. The Journal of Psychology, 77, 165-169.

Kudish, J. D. & Hoffman, B. J. (2002, October). Examining the relationship between assessment center final dimension ratings and external measures of cognitive ability and personality. Paper presented at the 30th International Congress on Assessment Center Methods, Pittsburgh, PA.

McCammon, S.L., Golden, J., & Wuensch, K.L. (1988). Predicting course performance in freshman and sophomore physics courses: Women are more predictable than men. Journal of Research in Science teaching, 25, 501-510.

McDaniel, M.A., Whetzel, D.L., Schmidt, F.L. & Maurer, S. (1994). The validity of employment interviews: a comprehensive review and meta-analysis. Journal of Applied Psychology, 79, 500-616.

McMurray, M.A., Beisenherz, F., & Thompson, B. (1991). Reliability and concurrent validity of a measure of critical thinking skills in biology. Journal of Research in Science Teaching, 28, 183-191.

Schmidt, F., & Hunter, J. (2004). General mental ability in the world of work. Journal of Personality and Social Psychology, 86, 163-173.

Spector, P. A., Schneider, J. R., Vance, C. A., & Hezlett, S. A. (2000). The relation of cognitive ability and personality traits to assessment center performance. Journal of Applied Social Psychology, 30(7), 1474-1491.

Steward, R.J., & Al Abdulla, Y. (1989). An examination of the relationship between critical thinking and academic success on a university campus. (ERIC Document Reproduction Service No. D 318936).

Watson, G.B., & Glaser, E.M. (1964). Watson-Glaser Critical Thinking Appraisal Manual. New York: Harcourt Brace Jovanovich

Watson, G.B., & Glaser, E.M. (1980). Watson-Glaser Critical Thinking Appraisal Manual. Forms A and B. The Psychological Corporation: San Antonio.

Watson, G.B., & Glaser, E.M. (1991). Watson-Glaser Critical Thinking Appraisal Manual: British Manual: Forms A, B and C. The Psychological Corporation: Harcourt Brace

Watson, G.B., & Glaser, E.M. (1994). Watson-Glaser Critical Thinking Appraisal Manual. Forms S. Harcourt Brace Jovanovitch: New York.

Watson, G.B., & Glaser, E.M. (2002). Watson-Glaser Critical Thinking Appraisal Manual. The Psychological Corporation: Harcourt Brace

Williams, R.L. (2003). Critical thinking as a predictor and outcome measure in a large undergraduate educational psychology course. (Report no. TM-035-016). Knoxville, TN: University of Tennessee. (ERIC Document Reproduction Service No. ED478075).

Wilson, D.R., & Wagner, E.E. (1981). The Watson and Glaser Critical Thinking Appraisal as a predictor of performance in a critical thinking course. Educational and Psychological Measurement, 41, 1319-1322.

Wood, L.E. (1981). An 'intelligent' program to teach logical thinking skills. Behaviour Research Methods and Instrumentation, 12, 256-258.

Miller, D.A., Sadler, J.Z., & Mohl, P.C. (1993). Critical thinking in preclinical course examinations. Academic Medicine, 68, 303-305.

Molidor, J., Elstein, A., & King, L. (1978). Assessment of problem solving skills as a screen for medical school admissions. Michigan State University, Report No. TM 800383. East Lansing: National Fund for Medical Education. (ERIC Document reproduction Service No. Ed 190595).

Morse, H., & McCune, G. (1957). Selected items for the testing of study skills and critical thinking. (3rd ed). Washington, D.C.: National Council for the Social Studies.

Mosher, J.P. (1999). The relationship between personality and management behaviour. Dissertation Abstracts International, 59, 6075.

Moutafi, J., Furnham, A., & Crump, J. (2003). Demographic and personality Predictors of Intelligence: A Study Using the Neo Personality Inventory and the Myers-Briggs Type Indicator. European Journal of Personality, 17, 79-94.

Ones, D.S., Viswesveren, C.J., Schmidt, F.L. (1993). Comprehensive meta-analysis of integrity test validities: findings and implications for personnel selection and theories of job performance. Journal of Applied Psychology, 78, 679-703.

Paul, R.W., & Nosich, G.M. (1992). A model for the assessment of higher order thinking. Paper commissioned by the United States department of Education, Office of Educational research and Improvement of the National Center for Education Statistics.

Robertson, I.T., & Smith, M. (2001). Personnel Selection. Journal of Occupational and Organisational Psychology, 74 (4), 441-472.

Rust, J. (2006). Advanced Numerical Reasoning Appraisal Manual. San Antonio, TX: Harcourt Assessment, Inc.

Sandor, M.K., Clark, M., Campbell, D., Rains, A.P., and Cascio, R. (1998). Evaluating Critical thinking Skills in a Scenario-Based Community Health Course. Journal of Community Health Nursing, 15 (1), 21-29.

Salgado, J., Anderson, N., Moscoso, S., Bertua, C., & de Fruyt, F. (2003). International validity generalisation of GMA and cognitive abilities. Personnel Psychology, 56, 573-605.

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.