**STEC Guidance: Raven’s 2 Progressive Matrices**

|  |  |
| --- | --- |
| **PURPOSE OF REVIEW** | To ascertain whether the test is appropriate for inclusion in the list of Suitable Tests for the Assessment of Specific Learning Difficulties (SpLD) |
| **DATE OF REVIEW** | 06/05/2020 |

|  |  |
| --- | --- |
| **GENERAL INFORMATION** | |
| Test | **Raven’s 2 Progressive Matrices** European Clinical Edition, Pearson Clinical |
| Publishers | Pearson Education Inc. (2018) |
| UK/US/Other | European norms |
| Date of standardisation /  re-standardisation | Earlier versions were published as Raven’s Progressive Matrices, including Raven’s Coloured Progressive Matrices (1947), Raven’s Standard Progressive Matrices (1938) and Raven’s Advanced Progressive Matrices (1962). From 1967-1995, many norm validation and regional standardisation studies were conducted. |
| Age range | 4.0- 69.11 |
| Open/Closed | Open to all with appropriate professional training. |

|  |  |
| --- | --- |
| **PRACTICAL CONSIDERATIONS** | |
| Apparent ease or difficulty of administration and scoring, from reading through the instructions. | **Area Assessed**  The Raven’s 2 isa **non-verbal test of general cognitive ability/intelligence** or ‘***g***’ as it is frequently termed. Specifically, the test claims to capture ‘**eductive’ ability**…the **ability to think clearly and solve complex problems**. In a somewhat convoluted theoretical statement, the authors are keen to point out that eductive ability is not the same as ‘*g*’ **BUT** the Raven’s tests do appear to be one of the best single measures of general intelligence available. However, ‘*g*’ should not be confused with *general academic ability* and it should not be thought that *g* encompasses all the abilities necessary for intelligent behaviour, which also include judgement and a store of specialist information.  **Administration**  The test is available in a digital and a paper form.  If the **digital version** is used this must be via the assessor’s computer(s) but the test can be administered to individuals or groups, provided enough computers are available. This test is **NOT** suitable for use via remote online video platform as the individually generated item bank cannot be shared via a screen sharing platform. For the digital version UK assessors need to register with Pearson’s Q-Global platform.  There is a **short** and a **long form** of the digital version. The short form takes approx. 20 mins, the long form a maximum of 45 mins, although discontinue rules can shorten the time taken on the test (after 6 incorrect responses the program terminates the assessment automatically) and for children aged 4-8 there is a 30 minute time- limit.  It is straightforward to use the **paper version**, although this is available only for individual administration. Scoring can be by hand (very easy as a template is provided) or Q-Global can be used for scoring and reporting by entering item responses into the program. Time limits on the paper form are 30 mins for ages 4.0-8:11 and 45 mins for ages 9.0-69.11. There are no discontinue rules on the paper form….examinees complete all 36 or 48 items.  **Scoring**  Scoring is very straightforward, whether undertaken digitally or manually. |
| Alternate forms? | Only in the sense that for the digital test, each individual receives a **unique randomly generated set of items.** |

|  |  |
| --- | --- |
| **NORMS AND STANDARDISATION** |  |
| 1. Availability of standard, scaled or other scores for each subtest. 2. Do the norms follow a normal distribution or are there  * big gaps between scores * low ceiling scores | Standardised scores by age-group are available, as are conversion tables for percentile rank, normal curve equivalents, stanines, descriptive classifications and age equivalents. Confidence intervals can be derived.  Norms follow a normal distribution across the age groups and SSs range from 40-160. |
| Type of standardisation sample (delete as appropriate) |  |
| Size and spread of Standardisation Sample: (*Also consider representation of age range, socio-economic backgrounds, ethnicity, gender)*  ***Green:***  *>500 (General Population)*  *>100 overall and >50 (SpLD Specific)* | **Standardisation**  Initial development in the US, where in 2016 there was a pilot study of newly created test items on a sample of 1307 individuals aged 4.0-90.11. This was followed by a standardisation stage with a sample of 1200 individuals collected from France, Germany, Netherlands, Flanders, Scandinavia, Spain and UK during 2018-9. Sample is stratified using census data by age, sex, race/ethnicity/education level/geographic region.  Issues:   * **21 different age bands** mean that the 1200 normative sample contains only 53-67 cases in each age band. * **Some of the reliability and validity data** was derived only from the US sample, not the new European sample. |
| If the test is based on a SpLD Specific norm, how was the presence of SpLD identified in SpLD sample group? | A proportion of children with special education and gifted classifications were included in the normative sample -approx 1.5% in each group. |

|  |  |
| --- | --- |
| **CONTENT VALIDITY** | |
| There are no sub-tests…this is one test with progressively more difficult items. | Do the test items adequately represent and relate to the trait of function that is being measured?  All of the Raven’s 2 items were newly developed, retaining the underlying cognitive constructs of previous versions of the test. Alongside the three traditional item types, I new item type was included. All items were designed with one or more colours. A new item bank of 329 items was created. If used digitally, an algorithm randomly selects items and assembles a unique test for each individual, ensuring equivalency in item content and reliability. This increases test security.  The authors claim that internal and external expert item reviews were conducted at each research stage to ensure adequate content coverage and relevance. Efforts were made to eliminate factors that might provoke irrelevant items response e.g. visual acuity, design complexity, colour usage, distractors, and unnecessary working memory load. The introduction of the discontinue rule of the digital form reduces random guessing, fatigue, time pressure.  Solving the Raven’s 2 involves several domains of cognitive functioning such as **perception, attention, inductive reasoning, classification and spatial ability, simultaneous processing, fluid reasoning, working memory**. Together these are described as elements of **eductive ability.** |
| **CONSTRUCT/CONCURRENT VALIDITY** | |
| Has the test been compared to other similar tests? If so, which ones, and what is the correlation between them:  **Red**: <0.60  **Orange/Green**: >0.60 | The Raven’s 2 was compared to its predecessors (CPM, SPM, APM), and with the NNAT3, KBIT-2 and WRAT5 using non clinical samples in the US standardisation study.  Coefficients varied: They were strongest, as might be expected, for the CPM, SPM, APM, NNAT3 and KBIT2 Non Verbal test across all forms (digital long and short, paper) of the test (all .71+). They were weakest for WRAT5 Spelling, Word Reading and Sentence Comprehension (.43 - .62).  There were two additional special group studies comparing the results of the gifted (n =174), and intellectually disabled (n =173) sub-groups with a matched control group. The gifted cohort scored **significantly higher** on the Raven’s 2 than matched control groups and the ID group scored **significantly lower**, supporting the use of this test as a valid instrument to assess the abilities of those who are gifted or intellectually disabled. |
| **RELIABILITY** | |
| Has the author conducted any repeat studies to establish the reliability of the test? (number of studies conducted, and size of samples for each study)  **Green:** Multiple studies | Reliability coefficients were obtained by using a split-half measure i.e. examining the correlation between the total scores of the two half-tests: reliability coefficients were moderate to high, depending on age: all over .75. with those for age groups 9+, all over .85  Reliability coefficients for the randomly generated item banks on the digital form of the test were also calculated using a method called Item Response Theory (IRT) marginal reliability. Average marginal reliability coefficients of the digital long form are either good (.80s) or excellent ().90s) across all age groups. The average marginal reliability coefficients of the digital short form are good, with an average of .80.  Reliability coefficients for the special group samples (gifted, intellectually disabled) were also calculated using split-half method…and found to be .85-.90. |
| Reliability data  **Orange**: reliability >0.75  **Green:** reliability >0.85 | 1. The Test-Retest stability data is taken from the US standardisation study…the Raven’s 2 was administered twice to 239 individuals, with test-retest intervals ranging from 21-64 days. Reliability coefficients all .81+ |
| **SEM DATA, CONFIDENCE INTERVALS** | SEMs provided and confidence intervals easily calculated. |

|  |  |
| --- | --- |
| Points to consider: | * Requires minimal verbal instruction and no spoken or written responses, thus minimising (but not eliminating) the impact that language skills and cultural backgrounds may have on test performance. * Can be used to assess through the full range of cognitive ability and across a wide age–range. * Digital testing enhances test security by allowing each examinee to take a unique set of items randomly selected from a large item bank. Assessors who use Q-Global can generate scores and reports very quickly. * Manual scoring and administration is very straight-forward. * A measure of non-verbal ability, this test claims also to be a good ***general*** indicator of intelligence (but not academic ability). * The test is not designed to identify an individual with a specific learning difficulty, although it is valid for assessing individuals with an intellectual disability or those in a gifted category. * The paper form of the test for adults lasts 45 minutes and can be rather arduous to complete. Adults will quickly realise that test items are getting progressively more difficult and unless positively oriented towards this type of task, may quickly feel discouraged, as once the instructions are given, the assessor has no further input until the task is complete. The short form of the digital test is slightly less reliable than the long digital form but still has reasonably good reliability and, as part of a wider assessment session, might be more accessible to the individual tested. * The test is explicitly **not suitable** for remote assessment via video-conferencing platforms. |