

## Does a test have to be normed for the country in which it is used?

Some questions and answers concerning the use of Lucid's tests outside the UK

### **Q1: Does it matter if I use a test with UK norms in non-UK countries?**

#### **A1: It depends!**

It is commonly accepted professional practice amongst psychologists and educational professionals around the world, to use and rely upon tests that do not have norms for the countries in which they are used. [If you are not sure about what norms are and how they are created, see the footnote at the bottom of this page.]<sup>1</sup>

Proper standardisation and norming of a test is a large and costly undertaking and for various reasons it is impractical to develop separate norms for each and every country that a test might be used in. Although most tests in use in the UK have been normed for the UK, nevertheless there are many examples of tests widely used and accepted by professionals and advisory and regulatory bodies in the UK that do not have UK norms (see Table 2 below). A professional judgement has to be made about whether a given test is suitable for use in a different country to that in which it was normed and/or to assess individuals who are linguistically, educationally or culturally different to the population on which it was standardised and normed. This judgement depends on several key factors, which are explained in *Q3: How will I know if a test is appropriate for my country?* below.

### **Q2: Are Lucid's tests being successfully used outside of the UK?**

#### **A2: Yes.**

For many years there have been a large number of users of Lucid's tests in countries other than the UK, including Australia, New Zealand, United States, Canada, South Africa, Malaysia, Singapore and Hong Kong. In addition, numerous British and International Schools all over the world use Lucid's programs. Many of the students being assessed in British and International Schools around the world are clearly not significantly different in language, education or culture to students living in the UK, and so there are no concerns about the appropriateness of Lucid's tests under these circumstances. In other cases, however, there may be some significant differences in language, education or culture between the students being assessed and students living in the UK. Nevertheless, users in all these categories have been very satisfied that Lucid's tests are valid and reliable, and have found that the

---

<sup>1</sup> Before a test is released, it is usually subjected to a process of statistical **standardisation** in which the test is administered to a large number of individuals across the age range for which the test was designed. Factors such as gender, general ability and socio-economic class should be balanced across the sample to avoid bias and ensure that, as far as possible, the sample is representative of the whole population. The standardisation data are then analysed statistically to create **norms**, which enable **raw scores** to be transformed into **standardised scores**. Standardised scores are independent of age and enable a direct comparison between the individual's performance and that of his or her peers across the population. Since standardisation and norming of a test is a major (and costly) undertaking, commercial considerations about whether to standardise and norm a test for any given country.

results they get can be trusted when applied to the understanding of student's learning problems and for making educational decisions.

Confidence in using Lucid's programs overseas can be built upon even further when independent research has shown that not only can it be successfully used with UK norms but that it has been shown to be valid for those who have English as a second language. For example, independent peer-reviewed research by Brooks et al (2011)<sup>2</sup> has shown Lucid's dyslexia screening system Lucid Rapid to be a valid and useful tool for screening Singaporean students for whom English is a second language. Lucid Rapid has three tests that are taken from either Lucid CoPS or LASS depending upon age, therefore this study also provides cross-cultural validation evidence for CoPS and LASS.

### **Q3: How will I know if a test is appropriate for my country?**

#### **A3: Ultimately, this is an empirical question.**

In the absence of good scientific findings, you should take into account certain key factors and make your own professional judgement. See Key factors below for guidance on this.

#### *Key factors in judging whether a test is suitable for your country*

There are three main dimensions to determining whether a test is likely to be suitable or not:

- (i) **The nature of the ability being tested** (e.g. language, literacy, maths, cognition), including whether it is a high- or low-level skill.
- (ii) **The degree of similarity** between the people in the country in which the test was normed and those in the country where it is being used, including factors such as language, education and culture (see below for further discussion of this).
- (iii) **The method and purpose of testing**, e.g. does the test rely on spoken language or is it a paper-based test that depends on reading and writing? Is the test delivered by a computer? Is the purpose of the test to assess levels of functioning or to screen for some difficulty?

(i) & (ii) – Table 1 below will help you make a judgement as to whether a UK test is likely to be valid and appropriate for assessing individuals in your country. Note, however, that Table 1 takes no account of the third dimension (iii), i.e. the method and purpose of testing, which is dealt with separately below.

(ii) – The judgement regarding 'similarity' between countries takes into account a variety of factors, including language, education and culture. The list below is intended to serve as a rough guide. However, it should also be borne in mind that many countries are undergoing rapid changes in education and culture towards a more 'western' life style, and that rampant globalisation of the English language (particularly affecting young people and those who aspire to more affluent life styles) will affect any judgement about 'similarity'. Within many countries there are substantial sections of the population (especially in rapidly-developing cities) where the degree of similarity to the UK is greater than in the rest of the country (e.g. Shanghai and some other cities in China, Rio De Janeiro and some other cities in South America).

---

<sup>2</sup> Brooks, G., Ng, V., Lim, B.H., Tan, W.P. & Lukito, N. (2011) The computer-based Lucid Rapid Dyslexia Screening for the identification of children at risk of dyslexia: A Singapore study. *Educational and Child Psychology*, 28, 33-51.

**Some examples of the ‘similarity to the UK’ categorisation for Table 1**

- **Very similar:** Ireland, USA, Canada, Australia, New Zealand – ie those that have *essentially the same* language, education and culture.
- **Fairly similar:** Most European countries (except Ireland), Russia, Israel, India, South Africa and most South-East Asian countries – ie those that *share several features* in language, education or culture (but not all three).
- **Rather different:** Japan, countries of the Middle-East (except Israel), Africa (except South Africa), South and Central America – ie those that *share only a few features* in language, education or culture.
- **Very different:** China, Central Asia – ie those that *share virtually no features* in language, education or culture.

**Table 1** Relationship between ability tested and degree of similarity of the country of the testee to the UK. **The tick indicates that the test is more likely to be appropriate.**

| Degree of similarity between the country and the UK | Language and literacy                          |   | Cognition                                      |   |
|---|--|---|--|---|
|   | High level skills<br>(e.g. reading & spelling) | Low level skills<br>(e.g. auditory discrimination & phonological awareness) | High level skills<br>(e.g. memory & reasoning) | Low level skills<br>(e.g. visual processing & motor skills) |
| <b>Very similar</b>                                 | ✓  | ✓   | ✓  | ✓   |
| <b>Fairly similar</b>                               |  | ✓   | ✓  | ✓   |
| <b>Rather different</b>                             |  |   | ✓  | ✓   |
| <b>Very different</b>                               |  |   |  | ✓   |

(iii) The method and purpose of testing

***How should the method of testing be taken into account?***

Broadly speaking, the more a test relies on reading and writing, the *less appropriate* it is likely to be for use in countries with fewer similarities (e.g. most paper-based tests of verbal reasoning). Conversely, tests that rely less on reading and writing are likely to be *more appropriate* for use in countries with fewer similarities (e.g. some paper-based tests of nonverbal reasoning). Since many computerised tests (e.g. Lucid’s tests) do not depend on reading and writing skills they tend to be particularly useful in countries with few similarities to the UK. Tests delivered individually by an assessor using the medium of spoken language may be considered intermediate between paper-based tests that depend heavily on reading and writing and computer-based tests.

***How should the purpose of testing be taken into account?***

It is important to distinguish between **assessment** and **screening**. An assessment will be subject to the standardisation and norming procedures described in the footnote on page 1. Assessment tests can be used for various purposes, such as diagnosis or monitoring progress. Assessment tests can also be used for screening but when used in this way, the teacher or psychologist must generally decide (using the norms) where the cut-off separating the positive cases from the negative ones should be placed (e.g. bottom 10%, bottom 5%). This is distinct from a test that has been designed *specifically* for screening, which is developed in a rather different way because it has to distinguish between negative and positive cases in a manner than minimises both **false positives** ('false alarms') and **false negatives** (missed cases). A screening test should therefore be tuned or calibrated to achieve this specific end result. This is not a straightforward process, because simply moving the cut-off point on an assessment test will usually have the effect of reducing the incidence of one type of misclassification while simultaneously increasing the other (e.g. reducing false positives while increasing false negatives, or vice versa).

The answer to the question: '*Does it matter that tests have UK norms?*' therefore depends largely on the degree of difference between the general population of the UK and the population to which the testees belong. The greater the degree of difference, the increasing risk that the norms may give a misleading impression of those students' abilities. For example, several psychological and educational tests normed in the United States are currently in widespread use in the UK despite that lack of UK norms, including:

**Table 2** Widely used tests in the UK that do not have UK norms

| Test widely used in the UK   | Norms |
|--|-------|
| WRAT (Wide Range Achievement Tests) – tests of reading, spelling and maths | US    |
| WRIT (Wide Range Intelligence Test)  | US    |
| WRAML2 (Wide Range Assessment of Memory and Learning)                      | US    |
| TOWRE-2 (Test of Word Reading Efficiency)                                  | US    |
| CTOPP (Comprehensive Test of Phonological Processing)                      | US    |
| GORT-4 (Gray Oral Reading Test)  | US    |
| GSRT (Gray Silent Reading Test)  | US    |
| Spadafore Diagnostic Reading test – NO NORMS – CRITERION-REFERENCED TEST   | NA    |
| Test of Handwriting Skills   | US    |
| Test of Silent Contextual Reading Fluency                                  | US    |
| Test of Silent Word Reading Fluency  | US    |
| WRMT (Woodcock Reading Mastery Tests)                                      | US    |
| Woodcock-Johnson III Tests of Achievement                                  | US    |
| KBIT (Kaufmann Brief Intelligence Test)                                    | US    |
| NNAT (Naglieri Nonverbal Ability Test)                                     | US    |
| Peabody Picture Vocabulary Test  | US    |
| Lindamood Auditory   | US    |

These tests are regularly used in the UK in assessments for examination access arrangements<sup>3</sup> and have been approved for the assessment of dyslexic students in higher education applying for Disabled Students Allowances.<sup>4</sup> Within the professions carrying out such assessments it is generally believed that the differences between UK and US populations are too small to invalidate the use of these US tests to assess UK students.

However, when considering the suitability of any particular Lucid test, the user should take into account the ability being assessed by the test, and the degree of difference between the general population of the UK and the population to which the testees belong. The chief areas of likely difference are: **education, language and culture**. Across the developed world all children receive at least 10 years' education and standards are broadly similar hence differences attributable to education are mostly small, and UK norms will generally be perfectly satisfactory. However, literacy and maths can be exceptions because teaching approaches in these areas can vary. In less developed countries many children receive a more limited education (and some no schooling at all) and in such cases educational differences are likely to have a much bigger impact on test performance.

In tests where **high-level language skills** are assessed directly or indirectly (e.g. reading and spelling) language differences will obviously impact on the results and this must be taken into account when interpreting scores. For example, a 14-year-old student with only 2–3 years' experience of written and spoken English would not be expected to perform in reading and spelling at a level comparable to that of the average 14-year-old English student. In such cases, using **age equivalents** rather than standard scores will usually be more helpful. (For information on age equivalents, consult the Teacher Manual for the appropriate Lucid test, or see Lucid Fact Sheet 51.)

On the other hand, in tests that involve **lower-level language skills** (e.g. auditory discrimination or phonological awareness) language differences will have a much smaller effect on test results, although the degree of overlap between the phonology of the languages will also be a significant factor (e.g. English and French have a lot of sounds in common, while English and Mandarin have very few sounds in common). During the pre-school period the brain learns to distinguish (and to produce) the phonemes that it hears regularly; speech sounds that are heard only rarely will not be acquired naturally. (Unfamiliar phonemes can be learned later on as a second language, but with considerable effort, and never quite as well.)

For more information about Lucid or the developments or research please visit the Lucid web site [www.lucid-research.com](http://www.lucid-research.com). The Lucid staff can be contacted by email [info@lucid-research.com](mailto:info@lucid-research.com), telephone +44 (0)1482 882121 or fax +44 (0)1482 882911.

Please note that the information contained in this document is correct at time of going to press.

<sup>3</sup> See Jones, A. (ed.) (2011) *Dyslexia: Assessing the need for Access Arrangements during examinations. A practical guide* (4th edition). Patoss.

<sup>4</sup> The approved list of tests for DSA assessments may be found at <http://www.patoss-dyslexia.org/catalogue/patoss-dyslexia.org/mediastore/REVISED-guidelines-September-2011.pdf>