

UNDERSTANDING EXACT



Specialist Consultants

Dr Jonathan Ferrier & Dr Joanna Horne

Digital Literacy Assessment
for 13 to 18+ years

Copyright © 2025 GL Assessment Limited

Published by GL Education

1st Floor, Vantage London, Great West Road, London TW8 9AG

www.gl-assessment.co.uk

GL Assessment Limited is a GL Education Group Limited company

Typeset by Mike Connor Design & Illustration

All rights reserved. Except for the quotation of brief passages in reviews of criticisms, or the copying of short extracts for bona fide training or demonstrations, or where expressly provided for in this guide, no part of this guide may be reproduced, stored in a retrieval system or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior written permission of GL Assessment Limited.

Exact Software is the exclusive World Copyright of GL Assessment Limited. Any attempts to copy, modify, adapt, translate, or create derivative works based on this software or otherwise to infringe copyright of GL Assessment Limited will be pursued to the utmost in the Courts of Law.

8th edition, revised August 2025

Contents

1	Introduction	7
1.1	What is Exact?	7
1.1.1	Rationale for the tests in Exact	7
1.1.2	Why are the tests in Exact speeded?	8
1.2	Exact and Access Arrangements	9
1.2.1	JCQ regulations on access arrangements	9
1.2.2	Cambridge Assessment – exam access	9
1.2.3	Why Exact is suitable for access arrangements assessments	10
1.2.4	Limitations regarding access arrangements assessments	10
1.2.5	Assessment of processing speed	11
1.2.6	Who can administer Exact?	11
1.3	Standardisation and norms	11
1.4	Validity of Exact	13
1.4.1	Construct validity tests and results	13
1.4.2	Validation of the Exact typing to dictation test	14
1.5	Reliability of Exact	15
1.6	Online version of Exact: Equivalence study	16
1.7	Advantages of computerised tests	17
1.8	Accessing Exact through Testwise	18
2	Details of each test	19
2.1	Word recognition	19
2.2	Reading comprehension	20
2.3	Spelling	21
2.4	Handwriting to dictation	22

2.5	Typing to dictation	23
2.6	Guidelines for test administration	24
2.6.1	Trial run-through including how to exit during a test	24
2.6.2	Testing environment and equipment	24
2.6.3	Student preparation	25
2.6.4	Supervision	26
2.6.5	Giving encouragement, prompts and feedback	26
2.7	Assessing students under age 13	27
2.8	Assessing students aged 19 and older	27
2.9	Retesting and repeated assessment	28
2.9.1	Cautions regarding retesting and repeated assessment	28
2.9.2	Rate of working	28
2.9.3	Practice effects	29
2.9.4	Regression to the mean	30
2.9.5	What interval should be allowed before retesting?	30
2.10	Assessing students who have limited English	31
3	Understanding results	33
3.1	Types of report	33
3.1.1	Results profile	33
3.1.2	Cautionary warning regarding dubious reading speeds	33
3.1.3	Assessor's comments	34
3.1.4	Results breakdown	34
3.1.5	Checking the scores from the dictation tests	34
3.1.6	Itemised responses	35
3.2	Types of scores	35
3.2.1	Standard scores	41
3.2.2	Confidence intervals	41
3.2.3	Percentile scores	42
3.2.4	Age equivalents	42



4 Use of Exact when applying for access arrangements 43

4.1 JCQ regulations43

4.2 Guidelines on using Exact43

5 Appendices 44

5.1 Typed dictation texts for FORM A.44

5.2 Typed dictation texts for FORM B.45



1 Introduction

1.1 What is Exact?

Exact is a suite of computerised tests used to support applications for exam access arrangements for students aged 13–18+. The Exact suite comprises standardised tests of the following areas of attainment:

- Word recognition
- Reading comprehension and reading speed
- Spelling
- Typing to dictation
- Handwriting to dictation

Test administration is carried out entirely by the computer. Each test begins with spoken instructions and practice items. The total suite takes between 30–40 minutes. Full details of the tests in Exact, including guidelines on test administration, are given in Section 2. Results, based on nationally standardised norms, are available after the test and are given in standard score and percentile score formats within the age range 13:0–18:11, and age equivalents for the age range 6 to 18 years are also provided. Guidance on understanding results and interpreting reports are given in Sections 3 and 4.

Exact has been specifically designed to meet the need for a group of tests that assess whether examination candidates should have access arrangements, such as extra time or use of a reader or scribe in written examinations. They are particularly aimed at GCSE and A-level examinations and the requirements of the Joint Council for Qualifications (JCQ), which represents awarding bodies based in England, Wales and Northern Ireland, some of which offer qualifications to overseas centres. Assessors should note, however, that Exact does not provide ALL the evidence required by JCQ when applying for exam access arrangements. Indeed, there is no single test currently on the market that can provide all the information necessary for the full completion of JCQ Form 8, which depends on a range of specialist assessment skills as well as thorough familiarity with current JCQ regulations, and calls for information from various sources. Exact provides a substantial amount of the assessment information required for Form 8. Section 4 gives guidance on this.

Exact has two forms of equivalent difficulty – Form A and Form B. This allows for repeated assessment if desired, although this should be carefully planned and with due consideration to the impact of possible practice effects. For further information on retesting see Section 2.9.

1.1.1 Rationale for the tests in Exact

There are three distinct, but interrelated, skills that are required by fluent readers: phonics, rapid word recognition and comprehension. Phonics comprises the sub-skills of grapheme-phoneme decoding (used when reading) and phoneme-grapheme encoding (used when writing). For

most students, phonic skills have been mastered by the age of 13. However, to become efficient readers, as well as decoding skills students also need to acquire rapid word recognition.¹ We have therefore concluded that rapid word recognition and comprehension are the key skills to be assessed in secondary school students and in individuals above this level. In Exact we have designed tests for these key reading skills; we have also included measures of spelling, handwriting and typing, which are central to the requirement for students to be able to record their work and display their knowledge and abilities in examinations.

1.1.2 Why are the tests in Exact speeded?

All the tests in Exact are *speeded* – i.e. they are performed against time limits. There are good reasons for this. From age 13 onwards the underlying skills in reading and writing should be largely automatic so that the mental focus can mainly be on understanding what is read and on conveying clear meaning in writing. Unless individual words in text are read quickly and effortlessly, it is extremely difficult to retain morphological elements (words, phrases, sentences) in working memory so that the overall text can be understood.² Similarly, unless the mechanical production of written words (letter formation, spelling, organisation, layout) can be carried out quickly and effortlessly when writing, it is extremely difficult for the writer to retain in mind a clear idea of what they intended to get down on paper. Hence, untimed tests are likely to give a misleading impression of the capabilities of students in secondary school and beyond. In particular, when students with specific learning difficulties are placed in the situation of a timed examination, their literacy skills are likely to be much worse than would be predicted from untimed measures of those skills.³

Arguably, writing to dictation (as in the Exact handwriting to dictation test) provides a purer and more reliable measure of writing speed than free writing because it is uncontaminated by the student's ability to create ideas. Research has shown that free writing speed is influenced by the topic chosen, teacher and administrative factors, and the extent to which students want to (or have been encouraged to) produce a really good piece of writing.⁴

¹ Nation, K. & Snowling, M.J. (2004) Beyond phonological skills: broader language skills contribute to the development of reading. *Journal of Research in Reading*, 27(4), 342-356.

Perfetti, C.A. (1985) *Reading ability*. New York: Oxford University Press.

² Lyon, G.R. (1998) Why reading is not a natural process. *Educational Leadership*, 55(6), 14-18.

Perfetti, C.A. (1985) *Reading ability*. New York: Oxford University Press.

³ Lesaux, N.K., Pearson, M.R. & Siegel, L.S. (2006) The effects of timed and untimed testing conditions on the reading comprehension performance of adults with reading disabilities. *Reading and Writing*, 19(1), 21-48.

Runyan, M.K. (1991) The effect of extra time on reading comprehension scores for University students with and without learning disabilities. *Journal of Learning Disabilities*, 24(2), 104-108.

⁴ Ferrier, J., Horne, J. & Singleton, C. (2013) Factors affecting the speed of free writing. *Journal of Research in Special Educational Needs*, 13(1), 66-78.

1.2 Exact and Access Arrangements

1.2.1 JCQ regulations on access arrangements

Exact has been designed to meet the requirements for examination access arrangements, most notably those regulated by the Joint Council for Qualifications (JCQ) for GCSE and GCE A-level examinations. Schools and colleges may apply for access arrangements in examinations for students with learning difficulties or other disabilities, which may take the form of extra time to complete written examinations, rest breaks, use of a word processor, or, provision of a reader or a scribe. Exact is already widely used by many schools and examination centres for access applications, and is accepted by JCQ, although it does not provide ALL the evidence required by JCQ when applying for exam access arrangements (there is no single test currently on the market that can provide all the information necessary for this purpose).

The JCQ publishes 'Access Arrangements and Reasonable Adjustments' (AARA) each autumn. JCQ AARA is subject to revision each year and assessors should therefore check these regulations on an annual basis. Hard copies are sent to centres in September, or the booklet may be downloaded from <https://www.jcq.org.uk/exams-office/access-arrangements-and-special-consideration/regulations-and-guidance/>.

Patoss (the professional association of teachers of students with specific learning difficulties) publishes a guide to assessing the need for access arrangements⁵ (for further information see www.patoss-dyslexia.org). As with all computer delivered group tests, Patoss recommends that when assessing for exam access purposes, Exact should be administered carefully so that individual student responses can be observed and monitored.

Communicate-ed (www.communicate-ed.org.uk) offers training (online and face-to-face) and resources for professionals involved with access arrangements.

1.2.2 Cambridge Assessment – exam access

There are two types of additional arrangements available if you are using exams from Cambridge Assessment: access arrangements; and centre-delegated access arrangements.

Access Arrangements

When applying to CIE (Cambridge International Examinations) for additional arrangements for your students, you need to submit the Access Arrangements: Preparation – Form 1 along with evidence to demonstrate that the student has a 'history of need' that justifies the application. Many of our assessments and screeners, including Exact and CAT4, can be used in this supporting evidence.

You can find more information on CIE's website.

⁵ Jones, A. (Ed.) (2011) *Dyslexia: Assessing the need for Access Arrangements during Examinations: A Practical Guide* (4th Edition). Evesham, Worcs.: Patoss.

1.2.3 Why Exact is suitable for access arrangements assessments

In exams, students are under the pressure of strict time limits, which may pose particular problems for those with difficulties in handwriting, reading or spelling. Indeed, it is for this very reason that students with these difficulties are often allowed extra time by the awarding bodies. It has already been pointed out that literacy tests that are not speeded are unlikely to reflect properly the levels of literacy competence of individuals of secondary school age or older, particularly in situations such as examinations (see Section 1.1.2).

Consequently, all the tests in Exact include an element of time pressure in order to recreate that feature of exam conditions. Thus, in the spelling test there is ample time for students to type each word and correct a simple mistake but not enough time for them to try out a variety of different spellings. In the comprehension test, because dyslexic pupils may have to read and re-read questions a number of times in order to fully understand them, we have not only set a time limit on the whole test but have also included a measure of reading comprehension speed, relating to the time taken for the questions to be understood.

1.2.4 Limitations regarding access arrangements assessments

Administrators wishing to use Exact when assessing eligibility for access arrangements should note that, under current JCQ regulations, the Exact word recognition test is not acceptable as a measure of reading accuracy when applying for a *reader* because it is a *timed* test of single word reading (similar to TOWRE 2). Section 7.5.10 of the JCQ AARA 2018-19 specifies an *untimed* test of reading accuracy. It will therefore be necessary to supplement Exact results with the results of a suitable standardised *untimed* reading accuracy test, e.g. WRAT5. Nevertheless, validation studies (see Section 1.4) have confirmed that the Exact word recognition test is an excellent test of single word reading, with scores very consistently located between those of WRAT4 and TOWRE. This test is therefore useful to help 'paint a picture' of a student's disabilities as advocated by JCQ when applying for *extra time up to 25%* (JCQ AARA 2018-19, Sections 5.2.2 and 7.6.1).

The writing tests in Exact take the form of timed writing (both handwriting and typing) to dictation. A poor score in the handwriting test is clear evidence of slow writing speed, while a satisfactory score in the typing test provides evidence of adequate typing skills such that the student would be capable of using a word processor in examinations. A below average handwriting to dictation standard score can be used as evidence for 25% extra time. This might indicate that the physical act of writing is slow for a candidate. Qualitative analysis of the handwriting itself might also provide evidence of illegible writing. The writing to dictation cannot, however, provide evidence of underlying processing issues when considering what to write, or organising thoughts into coherent writing. For this an assessment of 'free writing' will be required, and assessors should have such an assessment in their battery of tests and assessments (for further information on assessing free writing consult the Patoss website: www.patoss-dyslexia.org or the Patoss guide *Assessing the need for Access Arrangements during Examinations: A Practical Guide*). The Exact dictation tests are useful in showing those students for whom a keyboard may be more appropriate than a scribe. This is particularly relevant to students wishing to go to university, where fluency on a keyboard is important and scribes are not readily available.

1.2.5 Assessment of processing speed

Exact is a suite of literacy tests and does not contain any cognitive tests. However, another product, Recall, assesses working memory and processing speed in the age range 7:0 to 16:11. Results from these tests are acceptable measures of cognitive processing when applying for exam access arrangements, provided the student is not older than the test ceiling of 16 years 11 months. For further information on Recall, visit the GL website.

1.2.6 Who can administer Exact?

It depends on the purpose of the assessment. When used for the purposes of general assessment, almost any competent adult can administer Exact with minimal training and by following guidance in the manual. So, it does not have to be a teacher who administers the tests – it could be a teaching assistant, for example. However, interpreting results from Exact requires professional educational skills and so should be left to a qualified teacher. For this reason, Exact is only available for purchase by schools, qualified teachers, other educational institutions and some other professionals connected with education (e.g. speech therapists or careers guidance advisors).

When used for the purposes of assessing eligibility for exam access arrangements, JCQ AARA specifies that the assessment must be carried out by a suitably qualified person. The head of centre must satisfy themselves that this person is competent to carry out such assessments (JCQ AARA 2018-19, Section 7.3). This person then takes responsibility for selecting appropriate tests, interpreting the results and making the recommendations for access arrangements.

These requirements apply whichever tests are used, whether Exact or any others.

1.3 Standardisation and norms

Technically, 'standardisation' is the process used in psychometric test development to create *norms* so that the performance of students of different ages can be represented by means of scores that are independent of age. However, the term 'standardised' is sometimes used in a non-technical sense to refer to the *consistent administration* of a test – i.e. test instructions and methods of administration are the same for all who take the test. Because this non-technical usage can be misleading (e.g. users may assume that a test has standardised norms when in fact it hasn't), we only use the terms 'standardisation' or 'standardised' in strict accordance with technical psychometric usage.

The most common normative scores are standard scores and percentile scores. Standard scores have a mean (average) of 100 and a standard deviation⁶ of 15. Percentile scores place individuals on a 'ladder' of attainment from 1 to 100 compared with the population of that age; e.g. a percentile score of 70 means that 70% of people would have lower raw scores and 30% would have higher raw scores. (For further information about standard scores and percentile scores see Section 3.2).

⁶ The standard deviation is the most common statistic for expressing variability in a set of scores and is calculated as the average amount by which the scores in the set deviate from the mean.

The standardisation sample for Exact Word Recognition, Spelling, Typing to Dictation, and Writing to Dictation comprised 1,171 students aged from 11:0–18:11. The students were drawn from 17 schools across the UK, selected to give a representative spread of types of school and socio-economic profiles. The basic results are shown in Table 1. Exact test results for the standardisation sample.

The standardisation sample for Exact Reading Comprehension comprised 3094 students aged from 13:0–18:11. The students were drawn from 22 schools across the UK, selected to give a representative spread of types of school and socio-economic profiles. The basic results are shown in Table 1. Exact test results for the standardisation sample.

It can be seen in Table 1. Exact test results for the standardisation sample, that Forms A and B are very similar in all tests.

All raw data, except those for the word recognition test, approximated to normal distributions (symmetrical bell-shaped curves), with skewness (the degree of asymmetry of the distribution) and kurtosis (the degree of flatness and peakedness of the distribution) below the critical threshold of 1.0. The distributions of raw scores for the word recognition test were negatively skewed – i.e. scores were found to bunch towards the upper end of the scale. For this particular test, therefore, the raw score was transformed to give the normal distribution that is necessary for satisfactory generation of standard scores. This statistical transformation, which took into account the speed of response, had minimal effect on the scores of students with a raw score below the mean but, as intended, had a somewhat greater effect on scores above the mean. For students who have below average word recognition, over 70% of the variance in transformed score is predicted by raw score – i.e. for these students the standard score produced by this test is largely a function of their reading accuracy rather than speed. For students whose word recognition is highly skilled, however, differences are largely a function of speed rather than accuracy.

Table 1. Exact test results for the standardisation sample

Test	Form A		Form B	
	Mean	SD	Mean	SD
Word recognition (total number correct)	52.17	8.81	52.09	8.19
Word recognition (transformed score)	24.57	6.12	25.28	6.46
Reading comprehension (total number correct)	24.40	5.79	22.10	5.85
Reading comprehension speed (wpm)	79.61	28.87	72.32	31.90
Spelling (total number correct)	34.40	9.30	36.03	8.37
Typing speed (wpm)	21.25	7.51	22.37	6.85
Handwriting speed (wpm)	20.88	5.18	21.29	4.02

SD= standard deviation.

Normative results in standard score and percentile score form are incorporated into the Exact program. The norms are provided in 1-month age bands from 13:0 to 18:11 years. Age equivalents were calculated for the age range 6:0 to 18:11 (over this age, age equivalents become meaningless). Age equivalents in the 6:0 to 12:11 years range were calculated by extrapolation.

1.4 Validity of Exact

Validation of a psychological or educational test is not the same thing as the psychometric standardisation of a test, nor should it be confused with the *reliability* of a test. 'Reliability' generally refers to the extent to which a test can be expected to give the same results when administered on different occasions or by a different administrator, or the extent to which the components of a test give consistent results (see Section 1.6). 'Validity' is a measure of the extent to which the test measures what it is supposed to measure (e.g. reading or spelling ability). Validity is usually established by comparing the test with some independent criterion or with a recognised test of the same ability. Inevitably, this raises the thorny issue of what is the 'gold standard' – i.e. which is the 'best' measure of any given ability against which all others should be compared? Professional opinions differ as to the merits of various tests, and consequently there are no generally agreed 'gold standards' for assessing reading, spelling and writing. Hence, the conventional method of establishing test validity is to show that a new test produces results that agree reasonably closely with well-established test(s) of the same ability.

1.4.1 Construct validity tests and results

In validating Exact, the following established tests were selected for comparison: TOWRE (Test of Word Recognition Efficiency) – a speeded test of recognition of real words and nonwords; WRAT4 (Wide Range Ability Tests) Reading and Spelling – untimed measures of single word reading and spelling accuracy; the Edinburgh Reading Test – a measure of reading comprehension ability; and the Hedderly Sentence Completion Test – a test of handwriting speed. Note that WRAT4 and TOWRE have US norms but are nevertheless widely used in assessments for exam access in the UK. Exact has undergone separate studies with different samples for the validation studies and the standardisation study. An independent validation study of Exact was carried out in 2010-11 by Dr Joanna Horne of the Psychology Department, University of Hull, in four different schools in different parts of Britain and involved a total of 103 students. The results showed that all the tests in Exact correlate significantly ($p < 0.01$) with equivalent conventional (pen and paper or individually administered) tests that are in regular use for exam access assessments, clearly evidencing the validity of the tests in Exact. The results are shown in Table 2. Construct validity results for the tests in Exact*. Note that Reading Comprehension and Reading Speed are not included because the Reading Comprehension subtest has been updated since this study was carried out.

Table 2. Construct validity results for the tests in Exact*

Exact test	Comparison test(s) and correlation values			
Word recognition	TOWRE Single Word Reading Efficiency r=0.80	TOWRE Phonemic Decoding Efficiency r=0.84	WRAT4 Reading r=0.70	Edinburgh Reading Test r=0.74
Spelling	WRAT4 Spelling r=0.91	WRAT4 Reading r=0.70	TOWRE Single Word Reading Efficiency r=0.76	TOWRE Phonemic Decoding Efficiency r=0.87
Handwriting speed	Hedderly Sentence Completion Test r=0.54	Exact typing speed r=0.48		

* All correlations are significant at the $p < 0.01$ level.

It should be noted that the differential correlations shown in the table follow a logical pattern. The Exact word recognition test correlated more highly with the TOWRE tests than with WRAT4 Reading. This is because the TOWRE tests are speeded tests (like Exact word recognition), while WRAT4 Reading is an untimed test. Exact spelling shows a very high correlation with WRAT4 Spelling – higher than with the various reading measures. (Note that, as might be expected, reading and spelling skills tend to be significantly related: the correlation between WRAT4 Reading and WRAT4 Spelling, for example, was found to be 0.70, the same value as between Exact spelling and WRAT4 Reading).

To give some idea of expected levels of correlation, the correlation values between WRAT4 Reading and the other comparison tests were as follows: TOWRE SWE 0.64; TOWRE PDE 0.77; Edinburgh RT 0.67. These values are, in fact, lower than the corresponding values for Exact word recognition, suggesting that Exact word recognition has somewhat better concurrent validity than WRAT4 Reading.

1.4.2 Validation of the Exact typing to dictation test

Exact Handwriting Speed is significantly correlated with the Hedderly Sentence Completion Test (a commonly used measure of writing speed). Since there are no comparable tests of typing speed, no validation figures are given for this component of Exact. However, an independent study of the writing and typing to dictation tests in Exact has been published, and this provides support for the validity of this test ⁷. This paper reports on two studies using computer-based dictation tasks for measuring speed of typing and handwriting.

⁷ Horne, J., Ferrier, J., Singleton, C. & Read, C. (2011) Computerised assessment of handwriting and typing speed. *Educational and Child Psychology*, 28(2), 52-66.

In the first study, 952 students, aged 11-17 years, attending 19 different secondary schools, hand wrote and typed passages dictated by a computer. For both handwriting and typing, a very high correlation was found between speed calculated by the computer and that calculated by a human assessor, establishing that computerised calculation is a reliable, as well as convenient and timesaving method of establishing writing speed. There were greater age-related gains in speed of typing compared with handwriting, and greater variation in typing skill than handwriting skill. However, almost half of students with slow handwriting (below standard score 85) were found to have average or better typing speeds.

In the second study, 55 students aged 13-14 were administered these tasks together with the Hedderly Sentence Completion Test of handwriting speed. Despite the clear differences between the two test formats, a reasonable level of agreement was found between them. Almost one-third of students with slow handwriting in the computer-based task had not previously been identified as having support needs but would potentially be disadvantaged in written examinations. By eliminating the 'thinking' time involved in free writing, computerised dictation tasks give 'purer' measures, which can reveal physical handwriting and/or typing problems. They also simulate examination requirements more closely than mechanical repetitive tests of writing speed, and should be particularly helpful in establishing whether students need access arrangements in examinations.

1.5 Reliability of Exact

'Reliability' generally refers to the extent to which a test can be expected to give the same results when administered on a different occasion (test-retest reliability) or by a different administrator (inter-rater reliability), or to which the components of a test give consistent results (internal consistency). Note that this is not the same as the validity of the test (see Section 1.4).

Table 3. Reliability coefficients (Cronbach's alpha) for four of the tests in Exact shows the coefficients of reliability for each of the Exact tests, calculated using Cronbach's alpha statistic, which is a measure of the internal test consistency. Reading Comprehension results are not included here because this subtest has been updated since this study was completed. Note that the reliability coefficients shown in the table are all high (around 0.9). These results show that the tests in Exact have excellent reliability.

Although test-retest reliability is frequently quoted in test manuals, this measure is problematic because students are likely to remember items and answers from the previous assessment, which results in confounding memory factors. However, since Exact comprises two parallel forms, these can be compared in a test-retest situation, which is arguably a more satisfactory method of checking the test reliability since the test content is different in the two forms. To achieve this, Exact reading comprehension and spelling test data were collected from a total of 373 students aged 11-16 attending a large secondary academy in South London. The test-retest correlation coefficients over a period of six months were: spelling 0.757, reading comprehension accuracy 0.614, reading comprehension speed 0.511, these results all being statistically significant at $p < 0.001$. (Word recognition and writing/typing to dictation were not tested in this project.)

Table 3. Reliability coefficients (Cronbach's alpha) for the tests in Exact

Test	Reliability
Word recognition overall score	0.93
Word recognition regular words score	0.97
Word recognition irregular words score	0.96
Spelling overall score	0.91
Spelling regular words score	0.94
Spelling irregular words score	0.88
Typing speed	0.93
Handwriting speed	0.93

1.6 Online version of Exact: Equivalence study

The current version of Exact runs online, rather than from a CD. There have also been some updates to the illustrations, audio and animations. A study was carried out between September and December 2018 to evaluate the equivalence of the old and new versions. Note that Reading Comprehension results are not listed because this subtest has been updated since the study took place.

A total of 146 students from six schools (n=5)/colleges (n=1) in the UK completed both the old (CD) and new (online) versions of Exact, with an interval of four to six weeks between test sessions. Each school/college collected data on just one form of the test (Form A or Form B – these were randomly allocated to schools), and test order (either the old or new version being delivered first) was randomly allocated to account for order effects. Students were selected based on their birth day of month (with each school randomly allocated a day of the month) to avoid any selection bias within the student sample.

Pupil population data for the school sample (based on the 2016/17 figures, which were the latest available at the time of the study) were compared to the national averages (for state-funded secondary schools, not including special schools, where available; for FSM and absence, the national figures include special schools as these are the only figures available). The school sample was found to be not significantly different from the national average on any of the pupil population measures: number on roll ($t=0.36$, $df=4$, $p>0.05$, NS); percentage of female students ($t=0.29$, $df=4$, $p>0.05$, NS); percentage of students eligible for free school meals at any time during the past six years ($t=1.37$, $df=4$, $p>0.05$, NS); percentage of pupils whose first language is not English ($t=1.99$, $df=4$, $p>0.05$, NS); percentage of pupils with a statement/EHCP ($t=0.05$, $df=4$, $p>0.05$, NS); and overall absence ($t=0.41$, $df=4$, $p>0.05$, NS).

Of the 146 students in the sample, 8.2% were eligible (at the start of the study) for free school meals. Twenty students in the sample (13.7%) were recorded as having a special education need/disability (SEND). In half of these cases, the specific type of SEND was not reported (6.8%) but, where it was stated, these included Specific Learning Difficulties (SpLD: 2.7%), Speech, Language and Communication Needs (SLCN: 2.7%) and Moderate Learning Difficulties (MLD: 1.4%). The majority of the sample (91.1%) were White, 6.8% were Asian and 2.1% were Black. With regard to language, 8.9% of the sample's first language was not English.

The resulting correlations between the old and new versions of Exact are given in Table 4 (note that 'n' varies between subtests as not all students completed both versions of each subtest). For all subtests, the correlations meet the required standard of 0.70, to adequately demonstrate equivalence (according to the European Federation of Psychologists' Associations' test review model). In all cases, the correlations are very highly significant ($p < 0.001$).

Table 4. Correlations between old (CD) and new (online) versions of Exact

Exact subtest	Correlation
Word recognition ^a	$r_s = .742, p < 0.001; n = 136$
Spelling ^a	$r_s = .886, p < 0.001; n = 138$
Typing to dictation ^a	$r_s = .726, p < 0.001; n = 119$
Handwriting to dictation ^a	$r_s = .761, p < 0.001; n = 113$

^a Data don't meet the assumptions required for parametric testing; correlational analysis utilises Spearman's 'rho'.

1.7 Advantages of computerised tests

One of the great advantages of a well-designed computer-based test is that it does not require any special expertise on the part of the administrator. This applies to all the tests in Exact, which can be administered by any competent adult (see Section 1.2.6). Provided headphones are used, they can also be administered and undertaken in a room where other activities are taking place, and no special directions to the students are required other than to tell the student(s) which of the tests should be attempted, along with an explanation of the importance of moving through the tests quickly and of thinking carefully about responses (see Section 2.6.3).

Computers also provide more precise measurement, especially when complex cognitive skills are being assessed. Tests are administered in an entirely consistent manner for all persons taking the test, which enhances reliability of measurement. Timings and presentation speeds can be controlled precisely. The subjective judgment of the administrator does not affect the test outcome as it can in conventional tests. Exact is largely self-administered and results are available immediately; both of these factors help to reduce administrative load and avoid time delays.

There is good evidence that most students prefer computer-based tests to conventional tests (whether paper-based group tests or administered 1:1 by a teacher). This is particularly the case for students with below average literacy skills, who are more likely to feel intimidated by assessments and be embarrassed by their performance. Computer-based tests have generally been found to be less threatening and less stressful, which helps to ensure more reliable results.⁸ There is also evidence that there is less gender bias in computer-based tests than in conventional tests, so there are good reasons to regard computer-based tests as fairer, as well as being more consistent and objective, than conventional tests.⁹

⁸ Singleton, C.H. (2001) Computer-based assessment in education. *Educational and Child Psychology*, 18, 58-74.
BDA (2005) Practical solutions to identifying dyslexia in juvenile offenders: Report of a joint project of the British Dyslexia Association and HM Young Offender Institution Wetherby, 2004-5. Reading: British Dyslexia Association.

⁹ Horne, J.K. (2007) Gender differences in computerised and conventional educational tests. *Journal of Computer Assisted Learning*, 23, 47-55.

1.8 Accessing Exact through Testwise

In order to access Exact through the Testwise platform, please follow this link: <https://support.gl-assessment.co.uk/knowledge-base/platforms/testwise>

This will guide you through:

- Student management
- Sitting creation
- Taking the test
- Generating reports

2 Details of each test

2.1 Word recognition

Word recognition is a speeded test of the ability to recognise individual real words out of context (see Section 1.1.2 for explanation of speeded tests). It depends on fluent and efficient reading accuracy. In each item the student is presented with six real words shown on a circular location space in the centre of the screen (a circular location space is used, rather than random locations across the screen, to minimise any possibility of the results being adversely affected by visual tracking, fine motor skills and/or eye-hand coordination difficulties). One of those words (the 'target') is spoken by the computer, and the task is to identify the target word by clicking on it using the computer mouse as quickly as possible. (It is essential that students doing this test on a laptop use a mouse rather than a touchpad, because use of the latter creates an unacceptable delay in response times.) A total of five seconds is allowed for a response, with an audible prompt being given after three seconds. The remaining five words (the 'distractors') were selected in order to maximise lexical and phonological similarity with the target word, e.g.:

Target word: century

Distractors: centre, sanctuary, centrally, scented, central

Each of the two forms of the word recognition test comprises a total of 60 items, with equal numbers of regular and irregular target words. Difficulty – in terms of the frequency of the target words in written English – has been balanced across the two forms. The test begins with two practice items.

Because the distribution of raw scores for this test tends to be negatively skewed – i.e. scores tend to bunch towards the upper end of the scale – the raw score has been transformed to give the normal (bell-shaped) distribution that is required for generating standard scores. This transformation takes into account the speed of response, and has minimal effect on the scores of students with a raw score below the mean, but a somewhat greater effect on scores above the mean. For students who have below average word recognition, over 70% of the variance in transformed score is predicted by raw score – i.e. for these students the standard score produced by this test is largely a function of their reading accuracy rather than speed. For students whose word recognition is highly skilled, however, differences are largely a function of speed rather than accuracy.

Results are provided for overall words (see Section 3.1.1), and for regular words and irregular words (see Section 3.1.4). In practice, students do not use a phonetic approach when recognising words at speed, and their scores are usually very similar for both phonetically regular and irregular words. The standard scores in this test correlate well with spelling ability, and a standard score below 85 is a useful pointer to literacy problems such as dyslexia.

2.2 Reading comprehension

Each of the two forms of the reading comprehension test comprises five passages of increasing length and difficulty. Each passage is accompanied by a number of multiple-choice questions that demand good literal and inferential reading comprehension skills. The characteristics of each passage are shown in Table 4.

Table 4. Passages in the reading comprehension test

Item no.	Type of passage	No. of questions	Form A		Form B	
			<i>Title</i>	<i>Length</i>	<i>Title</i>	<i>Length</i>
1	Bulleted list	5	Caring for Your Venus Flytrap	83 words	Dragon Fruit Facts	84 words
2	Short story	10	The Storm	91 words	Lost and Found	102 words
3	Advertisement	10	Belldale Adventures	178 words	Woodworking Convention	157 words
4	Factual report	10	Axolotl Adoption	204 words	Dictionary Dedication	198 words
5	Excerpt from classic literature	10	<i>Cranford</i> (Elizabeth Gaskell, 1853)	243 words	<i>A Tale of Two Cities</i> (Charles Dickens, 1859)	234 words

Each passage is still available when the questions are shown. This is to avoid students having to remember the text, which would confound reading comprehension with memory ability. Each question has three possible answers. Some questions have the answer 'Impossible to say', which should be selected when the answer cannot be determined or inferred from the information given in the passage.

Note that 'Impossible to say' is not equivalent to 'Don't know', which would indicate the student's inability to deduce the answer, rather than the inherent indeterminacy of the answer; this is explained in the practice phase of the test.

The test begins with a practice item, after which students are allowed a maximum of 10 minutes for the whole test, with no restrictions on the proportions of time spent on the five passages. A timer is visible throughout the duration of the test, so that students can see how long they have left to complete the test. Students are permitted to return to a previous passage if they wish.

Students should be encouraged to work swiftly but conscientiously, and to make a fair attempt at each question. The overall score is designed to reflect careful reading and considered understanding, so students who adopt the tactic of moving on if they cannot immediately determine the answer to a particular question are likely to score rather poorly. Students need to appreciate that they have to think about the questions (as in an examination). Some students may be inclined to automatically click 'Impossible to say' if they are unable immediately to find the answer to a question within the passage. Students should be made aware that the answer

can often be deduced or worked out from the information given in the passage, and that the 'Impossible to say' option should only be selected when there is insufficient information given in the passage to enable the answer to be deduced.

Students should also be made aware that although the reading comprehension test only lasts 10 minutes, they are not expected to finish it in that time and they should take enough time to get the earlier (easier) questions right. The later passages are quite difficult – the last passage, in particular, will be a significant challenge for students at higher education level – so it is much easier to amass points for answers in the early passages, and students should not rush these.

Results for reading comprehension accuracy and reading comprehension speed are given (see Section 3.1.1). The reading comprehension accuracy score is the total number of correct answers given. The program calculates reading comprehension speed by dividing the estimated number of words read in each of the five passages (including the questions) by the time taken to read each passage and answer the associated questions. 'Dubious' or 'aberrant' cases – i.e. students who simply select answers at random without reading or giving proper consideration to the passages – are flagged up by the program as statistical 'outliers', and drawn to the attention of the administrator. These cases tend to have suspiciously high reading comprehension *speed* but low reading comprehension *accuracy* (see Section 3.1.2 for further explanation of this).

2.3 Spelling

Spelling is a speeded test of the ability to spell regular and irregular words (see Section 1.1.2 for explanation of speeded tests). Each form comprises 30 regular words and 20 irregular words that are presented in order of difficulty (which was established by previous trials with large numbers of students). In each item the target word is spoken by the computer both in isolation and in the context of a sentence. An illustration associated with the sentence appears on the screen. The task is to type in the target word as quickly as possible. Note that students can begin typing in the target word as soon as they hear it and do not have to wait for the contextual sentence to be spoken.

The time allowed for each item is a function of the *length* of the word (3–14 letters), with a minimum of 9 seconds and a maximum of 31 seconds. An audible prompt is given 3 seconds before the allowed time is up. However, the time allowed does *not* increase as a function of the *difficulty* of the words. There are two reasons for this. The first reason is that if the time allowed *did* increase as a function of the difficulty of the words, this would reduce the effectiveness of the test to identify poor spellers. The second reason is that, at this age, spelling is mostly *automatised* – that is, through practice the student has learned how to spell the word without much conscious effort – and hence the time taken to produce a spelling is largely due to how many letters have to be written or typed. Hence the test is efficient in revealing the lack of automaticity in poor spellers (see Section 1.1.2 for a discussion of the importance of automaticity in skilled literacy).

The spelling test is *adaptive* insofar as the entry point is determined by performance on a number of 'probe' items of increasing difficulty given at the start of the test. When a student fails a probe item (or all the probes have been successfully answered) the program jumps to the appropriate place in the test. This preserves the sensitivity of the test for assessing poor spellers whilst avoiding boredom of more able students, who would otherwise find it very tedious and

demotivating to have to spell lots of very easy words. Items that are jumped in this way are credited as having been passed correctly by the student. The test is automatically discontinued when four out of the last six items attempted have been answered incorrectly. Since all the items are in order of difficulty the score obtained still accurately reflects the student's spelling ability even though they have not necessarily attempted all the items.

Results are provided for overall words (see Section 3.1.1), and for regular words and irregular words (see Section 3.1.4). A breakdown of all responses in this test is also provided for assessors who wish to use this information diagnostically. In secondary school students, spelling is often a more accurate measure of dyslexia than reading ability, and standard scores below 85 may be another pointer to that diagnosis.

2.4 Handwriting to dictation

There are a limited number of tests available for assessing the speed of handwriting, and the most widely used of these, which includes standardised norms, is an assessment of free writing. However, these tests include such a large and variable element of thinking time that they cannot offer a reliable measure of actual writing speed. We have therefore designed a handwriting to dictation test which eliminates thinking time and is a pure measure of handwriting speed and legibility. Inevitably, the handwriting speed for a dictated piece is very different from the speed obtained from free writing, and the two should not be confused. Both types of test may be required to fully assess a student's handwriting difficulties.

In the handwriting to dictation test a passage is dictated by the computer and the student has to write the dictated text by hand. A total of seven minutes is allowed for the passage, which is approximately 200 words in length. The passage is dictated in chunks of around four to six words, followed by a pause. The student has to press the 'page down' key to hear the next chunk or can press the 'control' key to hear that chunk repeated. In order to avoid confounding spelling skills with writing speed, the passage has been designed to impose minimal demands on spelling skills in the early paragraphs.

The results for this test show the speed of handwriting in words per minute (see Section 3.1.1), and the number of words handwritten (see Section 3.1.7). The number of handwritten words is estimated by the program based on the number of words dictated. Our research has shown that in 95% of cases the computer estimation of the number of words is sufficiently close to the actual number of words that it makes no appreciable difference to the standard score. However, in about 5% of cases, there may be substantial divergence between the computer estimation of the number of words and the actual count. This may arise because the student has been listening to the story and not writing.

Administrators should therefore always carry out a visual inspection of the handwritten work and, where anomalies are suspected, the actual count should be entered into the computer to replace the estimated count (see Section 3.1.5 for an explanation of how to do this).

In addition, if desired, the administrator can count up the number of misspelled or illegible words and enter these into the computer, and the program will calculate these measures as percentages of the total number of words typed or handwritten (see Section 3.1.4). See Section 3.1.5 for an

explanation of how to do this. However, it should be stressed that this is not intended to be a test of spelling and hence most students make few errors. In the standardisation sample, about 70% of students made fewer than 5% spelling errors in either passage.

2.5 Typing to dictation

Since it is a requirement by the examining authorities that students should be proficient on a keyboard in order to be allowed to use a word processor in examinations, we have included a typing to dictation test which can be compared with the handwriting to dictation score.

In the typing to dictation test a passage is dictated by the computer and the student has to type the dictated text using the computer keyboard. A total of seven minutes is allowed for the passage, which is approximately 200 words in length. The passage is dictated in chunks of around four to six words, followed by a pause. The student has to press the 'page down' key to hear the next chunk or can press the 'control' key to hear that chunk repeated. In order to avoid confounding spelling skills with typing speed, the passage has been designed to impose minimal demands on spelling skills in the early paragraphs.

The typed dictation is often of interest because it clearly demonstrates spelling difficulties and problems with auditory memory, both of which may suggest dyslexia.

The results for this test show the speed of typing in words per minute (see Section 3.1.1), and the number of words typed (see Section 3.1.4). The number of typed words is calculated by the program, which treats a string of characters separated by spaces as 'words' and counts up accordingly. Our research has shown that in 95% of cases the computer estimation of the number of words is sufficiently close to the actual number of words that it makes no appreciable difference to the standard score. However, in about 5% of cases, there may be substantial divergence between the computer estimation of the number of words and the actual count. This may arise because the student has neglected to put spaces between several words.

Administrators should therefore always carry out a visual inspection of the typed piece and, where anomalies are suspected, the actual count should be entered into the computer to replace the estimated count (see Section 3.1.5 for an explanation of how to do this).

In addition, if desired, the administrator can count up the number of misspelled or illegible words and enter these into the computer, and the program will calculate these measures as percentages of the total number of words typed (see Section 3.1.4). See Section 3.1.5 for an explanation of how to do this. However, it should be stressed that this is not intended to be a test of spelling and hence most students make few errors. In the standardisation sample, about 70% of students made fewer than 5% spelling errors in either passage.

2.6 Guidelines for test administration

2.6.1 Trial run-through including how to exit during a test

Assessing students with Exact is straightforward but before the teacher or administrator attempts to test any student it is advisable first to run through the complete suite of tests to familiarise themselves thoroughly. Ideally, this should be done in the same or similar test situation in which the students will be (see Section 2.6.3). To do this, register yourself as a student and generate an access code for yourself.

2.6.2 Testing environment and equipment

The ideal testing environment is one that is reasonably quiet, with minimal distractions. Ideally, this should be a separate room, but Exact has been designed to be robust for use in the ordinary classroom, provided visual and auditory distractions (both to the student being tested and to other students in the class) have been minimised. To minimise auditory distraction, headphones are recommended. Inexpensive lightweight headphones of the type used for portable audio equipment will be adequate (but not the type that are inserted into the ear). Teacher or supervisor judgment is paramount in ensuring the appropriate testing environment.

If assessment is going to be carried out in an ordinary classroom in which there are other pupils, the computer and the student should be positioned in such a way that the student is not looking directly at the rest of the class, nor should the rest of the class easily be able to see the monitor screen. The best position for this is usually in the corner of the room. Students should not attempt the tests when other students are in a position in which they can become involved in the task or act as a distraction. It would be hard for other students to inhibit their reactions, and their behaviour could influence the decisions of the student being tested.

The teacher or supervisor should check that the equipment being used for the assessment is functioning correctly. This includes checking:

- that the sound system (speakers or headphones) is audible (not too loud or too soft, and without interference); and
- that the mouse is functioning correctly (non-optical types, particularly, require regular cleaning) and is positioned in front of the student on a suitable surface so that its movements are unimpeded. Please note that Exact should be used with a mouse (wired or wireless), not a touchpad as this will affect response times.

Exact should not be used for testing when any other applications are running on the computer, as these can interfere with the timings and recording of results. Please close down all other applications before starting Exact.

2.6.3 Student preparation

Before testing, each student must be registered on Testwise. See Section 1.8 for guidance on this. The tests can be done in any order but it is usually best to start with word recognition which students generally find quick and easy. Instructions are spoken by the computer, and each test commences with a practice or demonstration of the task. When the student has completed the practice items, the test phase begins.

The student should be sitting comfortably at a suitable level in front of the computer screen (not too high or low, in order for them to see the screen and use the mouse satisfactorily). It is not recommended that students attempt the tests standing up, as they are more likely to move about and alter the angle at which the screen is viewed – this can lead to failure to see everything that is happening on the monitor and can also disrupt mouse control. The supervisor should check for reflections on the monitor from windows and lights that could impair the student's perception. To do this the supervisor should check by viewing the screen from the same position that the student will adopt.

If necessary, students should be shown how to indicate responses to the computer using the mouse, and when to respond (essentially when the tests will allow them to respond). This is particularly important when testing students with physical disabilities. As with any format assessment, students should not be allowed to take the tests if they are unwell, as results are likely to be unreliable.

Most students will experience no difficulties in understanding what is required of them when taking the tests in Exact, enabling them to follow the practice tasks easily and progress to the test phase without special attention from the teacher or supervisor. However, it is important that the administrator ensures that students understand the nature of the tasks in Exact: that they are tests and not games, and that they must work swiftly but thoughtfully and try their best at all times.

This is particularly important in the reading comprehension test, in which some students may be inclined automatically to click 'Impossible to say' if they are unable immediately to find the answer to a question within the passage. Students should be made aware that the answer can often be *deduced* or *worked out* from the information given in the passage, and that the 'Impossible to say' option should only be selected when there is insufficient information given in the passage to enable the answer to be deduced.

Students should also be made aware that although the reading comprehension test only lasts 10 minutes, they are not expected to finish it in that time and they should take enough time to get the earlier (easier) questions right.

In the rare event that a student does not understand the instructions spoken by the computer, the supervisor may re-express them in a more suitable manner. Explaining and re-expressing the task requirements to the student may continue into the demonstration and practice stages of each test. This is particularly useful for any student who is experiencing problems in understanding the true nature of the task. It is often easier for the student to comprehend the task requirements by experience of the practice stages than by more abstract oral explanation. Once the test items commence there should be no further aid given to the student.

2.6.4 Supervision

Students should be closely supervised while attempting the tests, particularly when assessing the need for exam access arrangements, where JCQ regulations require the assessor who signs the JCQ forms to carry out the assessment (see Section 1.2.6). As with all computer delivered group tests, Patoss recommends that when assessing for exam access purposes, Exact should be administrated carefully so that individual student responses can be observed and monitored. The tests in Exact have been designed to be interesting and stimulating for students in this age group and the vast majority of students are highly motivated to do their best. Once the teacher is satisfied that the student understands the requirements of a test, has completed the practice items and has moved on to the test items, the teacher may leave the student to complete that test.

Where the teacher suspects that a student may not be well motivated to complete the test, or may be easily distracted, or may be performing deliberately below their capabilities, closer supervision will be necessary. Disaffected students may display non-compliance by clicking on test items or answers at random, rather than thinking about the tasks and selecting answers after proper consideration. Such students, or those with very low ability, may need close supervision in order to provide encouragement and ensure they remain on task. This is particularly important in the reading comprehension test, which requires careful thought (see Sections 2.2 and 2.6.3).

In order for the assessment to be 'fair' (i.e. to give a reasonably accurate representation of the student's abilities) it is essential to ensure that during the test:

- the student is paying attention, is 'on task', is not distracted and is trying their best
- the student does not become unduly fatigued
- there is no teaching or helping with the task during the test items (whether from the supervisor or other students)
- feedback from the supervisor is minimised and encouragement consistent (see further comments below).

2.6.5 Giving encouragement, prompts and feedback

As much as possible, the supervisor should avoid giving specific feedback to students during a test, because this may influence their behaviour in an undesirable fashion. This is good practice in any testing situation. There is a risk of feedback differentially affecting students, so that some are encouraged and others discouraged. Nevertheless, some students (particularly younger children or children with special educational needs) will try to elicit feedback from the supervisor about their performance. This may take the form of both verbal and nonverbal behaviours. For example, the student may ask directly if they were correct. Many students will look for the supervisor's facial and bodily reactions to their responses. Some students may even try to evaluate the supervisor's reaction by observing the supervisor's reflection in the monitor screen. For these reasons it is usually preferable that if the supervisor is going to be near the student to observe the assessment, they should sit to the side and slightly behind the student to minimise any feedback to the student which may bias the results.

Rather than specific feedback, general encouragement should be given to the student. This encouragement should be referenced to task completion rather than task accuracy and ideally should be delivered equitably to all students. However, it is inevitable that some students will require more encouragement than others, and where this is the case the teacher should be mindful of the possibility of influencing results unduly. Differential encouragement between students is likely to have an influence on the results obtained and therefore should be avoided where possible. Some key phrases and general incentive prompts which may be used to aid the administration of the tests include: “well done”; “you were good at that; now try the next one”; “you will like this game”; “now concentrate on this”; “try hard”; “listen very carefully”; “have a go at these ones”; “have a try”; “just do your best”.

2.7 Assessing students under age 13

It is standard practice that normative tests are not generally recommended for use outside the age range for which they have been standardised. Any test, such as Exact, which meets basic psychometric criteria must be standardised on a given population, and this will determine the range of applicability of the test (see Section 1.4 for an explanation of the standardisation process.) Tests appropriate to the student’s chronological age should be used wherever possible, to avoid the danger of inappropriate decisions being made – e.g. that a student is ‘at risk’ (or not ‘at risk’) when the evidence for this may be unsound.

Administrators who require a test to discern a child’s reading skills should consider using NGRT (ages 6-16).

2.8 Assessing students aged 19 and older

Exact was designed for use with students aged up to 18 years 11 months, and use with students older than this can create uncertainties when interpreting results. If the student is older than 18:11 then the program will use the norms for 18-year-olds when analysing results. However, international research studies have shown that, for the vast majority of the population within developed countries, literacy skills do not alter significantly in adulthood.¹⁰ Consequently, it is acceptable to use Exact with adults aged 19 and over if there are no satisfactory alternative tests available. Normative results for adults aged 19-54 should not be significantly different to those for adults aged 18; nevertheless, administrators should exercise caution when drawing conclusions about results of older adults.

¹⁰ Satherley, P. & Lawes, E. (2008) The Adult Literacy and Life Skills (ALL) Survey: Age and Literacy. Ministry of Education, New Zealand.
 Scottish Government (2009) Scottish Survey of Adult Literacies: Report of Findings, Part 4. Edinburgh: The Scottish Government.
 Cascio, E., Clark, D. & Gordon, N. (2008) Education and the age profile of literacy into adulthood. (Working Paper 14073) Cambridge, MA: National Bureau of Economic Research.
 Sloat, E. & Willms, J.D. (2000) The International Adult Literacy Survey: Implications for Canadian Social Policy. Canadian Journal of Education, 25(3), 218-233.

2.9 Retesting and repeated assessment

Exact has two forms of equivalent difficulty: Form A and Form B; and they allow for retesting or repeated assessment if desired. Exact was designed primarily for identifying students who require access arrangements in examinations – i.e. for identifying significant weaknesses in literacy skills – not for *continuous assessment*, which is focused on measuring improvement in literacy skills as a result of educational input.

2.9.1 Cautions regarding retesting and repeated assessment

When embarking on retesting or continuous assessment, it is particularly important to remember that when students are assessed on any psychometric test (whether administered conventionally or by computer) and the test (or a parallel form of it) is given again some time later, it must not be expected that the scores of all students will either stay the same or increase. Inevitably some will show a decline in scores, but this should not be taken to indicate that these latter students have necessarily decreased in the relevant ability. The reasons for this include not only the unsystematic and unpredictable variations to which all human performance is naturally subject, but also certain systematic factors that can dramatically influence test results. These factors include: rate of working, practice effects and regression to the mean. If misinterpretation of results is to be avoided when Exact is used repeatedly, it is important that administrators understand these factors and are fully aware of their possible impact on results.

2.9.2 Rate of working

Since Exact was designed for access arrangements assessment, all the assessments in the suite have strict time constraints (i.e. they are ‘speeded’ tests). The reason behind this has already been explained in Section 1.1.2. In all speeded tests the rate at which the student works is an inherent factor in determining the results. The type of time constraint differs across the tests in Exact and this is more important in certain tests than others. In the tests of word recognition and spelling, a given time is allowed for each item; a fixed time limit (5 seconds per item) in the word recognition test and a variable time limit (geared to word length) in the spelling test. In the reading comprehension, handwriting to dictation and typing to dictation tests, an overall time limit is imposed rather than a time limit per item. In reading comprehension, 10 minutes is allowed for the whole test, and students are required to attempt as many items as they can within that time (but not necessarily to attempt all the items). Both handwriting to dictation and typing to dictation have an overall time limit of 7 minutes.

The word recognition and spelling tests may be regarded as *simple tests* because:

- they comprise a large number of items
- items are independent of each other
- the student either knows or does not know the answer
- as soon as one item has been completed the student is immediately presented with the next.

This means that the task is automatically paced by the time constraints placed on each item. If an item is not completed within the time limit, the program automatically advances to the next item, and so on. Although variation in speed of working between different students is likely to affect the results, variation in the speed at which an individual student works on different testing occasions is unlikely to affect the results very much. Consequently, time is a less important factor in these tests, and differences in scores on these tests from one occasion to the next are principally a result of changes in student ability rather than speed of working.

The dictation tests employ a different type of time constraint but the nature of the tasks means that each item (i.e. a phrase heard by the student) can be regarded as essentially independent of the other items (i.e. previous and subsequent phrases in the passage). As soon as the student signals they have finished writing or typing one phrase, the computer automatically gives the next phrase, and so on until the time limit is reached. The task is self-paced and, in principle, differences in scores on these tests from one occasion to the next could arise from the student not working as hard on one of the occasions as on the other. In practice, however, it turns out that time is not such an important factor because item independence coupled with task simplicity results in differences between testing occasions being principally attributable to changes in writing ability.

The reading comprehension test contrasts markedly with the other tests in the Exact suite. Text passages are presented, and several items (in the form of multiple-choice questions) relating to each passage have to be attempted within the time limit. In order to answer questions, the student may have to refer back to the text or consider answers to previous questions. Hence this test may be regarded as *complex* rather than simple because items within a passage are not independent and it is not simply a case of either knowing or not knowing the answer but, rather, of being prepared to devote sustained mental effort at an optimum rate over the whole task in order to work out each answer. The 'optimum rate' is a speed that is consistent with their word recognition and verbal comprehension ability. If they exceed their optimal speed, they will make more word recognition errors and be more likely to misunderstand sentences, which will result in a lower score. If they read slower than their optimal speed, they will have insufficient time to attempt all the passages and so be prevented from the opportunity of answering more questions than the less able readers, which will also result in a lower score. Either way, the student will appear to be a less able reader than is really the case.

It should be apparent that if the student is tired, less well-motivated or not in a positive mood, or if they perceive that the consequences of less effort will not matter very much, they will tend to work slower and be less inclined to put in the necessary cognitive effort. If this happens, their score will be unlikely to reflect their true ability.

2.9.3 Practice effects

'Practice effects' are the positive or negative psychological impacts of previous assessment(s) on a student's performance..¹¹ Positive impacts, which include factors such as item familiarity and increased confidence as a result of previous experience with the tasks, tend to inflate scores on subsequent assessment occasions. Negative impacts, which include factors such as decreased

¹¹ For further explanation of practice effects see: Kulik, J.A., Kulik, C-L.C. and Bangert, R.L. (1984) Effects of Practice on Aptitude and Achievement Test Scores. *American Educational Research Journal*, 21, 435-447.

motivation due to boredom with the tasks or overconfidence as a result of feedback from previous assessments, tend to deflate scores on subsequent assessment occasions. In general, the magnitude of practice effects is a function of how often students have been assessed and the time interval between assessments. Both positive and negative psychological impacts tend to increase as the time interval between assessments decreases. Furthermore, practice effects will not necessarily affect all students to the same extent. Some students may experience more negative effects, while others may experience more positive effects.

2.9.4 Regression to the mean

All test scores, by their very nature, are variable. On any psychometric test the actual score obtained is an estimate of the student's true score, which will fall within a certain range of the actual score; this range is known as the 'confidence interval'. It means that one can have a certain level of confidence (in this case 90% confidence) that the student's true score lies within a range of the actual score that is equal to plus or minus the 'confidence interval'. On another occasion on the same test, the same student is likely to score slightly differently, which could be higher or lower than the previous score. The confidence interval for any test is determined by the test's reliability – i.e. the extent to which it can be relied on to give the same result on another occasion.

Consequently, human beings do not perform at the same level on every occasion, and some assessment tasks are more influenced by this variability. Over time, a person's skills may increase as a result of learning, practice and general experience. However, many other things also influence performance, such as mood, motivation, tiredness, instructions and perceived consequences. As explained above, *simple tests*, i.e. ones where the student either knows or does not know the answer (e.g. Exact word recognition and Exact spelling) are less subject to such influences than more *complex tests*, such as Exact reading comprehension, where it is not a case of either knowing or not knowing the answers but, rather, of being prepared to devote sustained mental effort at an optimum rate over the whole task in order to work out the answers.

Coupled with the general tendency of any measurement process to involve a degree of random error, these natural variations in test scores result in the statistical phenomenon known as 'regression toward the mean', whereby a score that is extreme (meaning further away from average performance) on its first measurement, will tend to be closer to the average on its second measurement, and if it is extreme on its second measurement, it will tend to have been closer to the average on its first..¹²

2.9.5 What interval should be allowed before retesting?

The previous three subsections make clear that, when carrying out repeated testing, variation in test performance is always to be expected, and gains cannot be counted on. Even when the best teaching has been provided, it is likely that a few students will exhibit apparent drops in performance from one test occasion to the next. This is due to various factors, including the impact of rate of working (more noticeable in complex as opposed to simple test formats),

¹² Upton, G. & Cook, I. (2006) *Oxford Dictionary of Statistics*, Oxford University Press.

Stigler, S.M. (1997). Regression toward the mean. *Statistical Methods in Medical Research*, 6, 103-114.

Tweney, R.D. (2013) Reflections on regression toward the mean. *Theory and Psychology*, 23, 271-274.

practice effects (more pronounced if the interval between testing occasions is short), and regression toward the mean (scores that are extreme on the first measurement will tend to be closer to the average on the second measurement). When interpreting results and conveying results to pupils, teachers or parents, it is essential that administrators take these factors into account and avoid drawing naïve or simplistic conclusions from changes in scores from one testing occasion to the next.

Therefore, it is recommended that, in normal circumstances, the interval between successive assessments should be preferably one year or, at the very least, one term or semester. Even though there are two parallel forms, if the period between successive assessments is relatively short (i.e. a matter of weeks or up to a school term or semester), practice effects could still arise and confound results. Research has shown that when retesting after a long school holiday performance is more likely to have declined..¹³

Occasionally exceptional situations may arise where a teacher needs to re-administer one or more of the tests in Exact after a much shorter interval, e.g. if it is discovered that when the student first took the tests he or she was unwell, or where a fire drill interrupted the assessment, or if the student was clearly not applying proper attention or effort to the tasks. In such cases, the results are unlikely to give a true indication of abilities and it is permissible to re-test the student. Nevertheless, there should be a delay of at least two weeks before re-administering the test(s) and the alternative form should be used. The first result should be discarded and the second result should be regarded as the true result.

2.10 Assessing students who have limited English

Assessment of any student who has limited proficiency in spoken or written English is often problematic. All the tests in Exact require considerable knowledge of written English and students who lack this knowledge would be expected to be impaired across the full range of measures in the suite. Exact results of students for whom English is an additional language should therefore be considered in relation to the level of English knowledge of the student, with the conclusion being modified in the light of this. Factors that should be taken into consideration include whether or not English is one of the languages spoken in the student's home, how long the student has been living in an English-speaking environment, and how long the student has been educated in English.

When using the results of Exact in conjunction with applications for exam access arrangements, it is important to be aware that limited proficiency in spoken or written English solely because English is not the student's first or main language (EAL) is not a criterion for having difficulties in learning, cognition, language or communication. What must be evidenced is that: "The candidate must have an impairment in their first language which has a substantial and long-term adverse effect. A candidate does not have a learning difficulty simply because their first language is not English, Irish or Welsh." (JCQ AARA 2018-19, Section 4.1.2). Consequently, where it is suspected

¹³ Sainsbury, M., Whetton, C., Mason, K. and Schagena, I. (1998) Fallback in attainment on transfer at age 11: evidence from the Summer Literacy Schools evaluation. *Educational Research*, 40, 73-81.

Davies, B. & Kerry, T. (1999) Improving student learning through calendar change. *School Leadership and Management*, 19(3), 359-371.

that an EAL candidate has difficulties that might qualify for exam access arrangements, great care is taken to evidence this by means of appropriate test results that show impairment in cognitive skills such as working memory, processing speed, etc., and by showing that the student's experience of English has been adequate enough for normal literacy skills to be expected.

For further information on assessment of learning difficulties in literacy (including dyslexia) in EAL students and other multilingual students, see Cline (2000), Cline and Frederickson (1999), Cline and Shamsi (2000), Durkin (2000), Gunderson, D'Silva and Chen (2011), Peer and Reid (2016), and Tsagari and Spanoudis (2013).¹⁴

¹⁴ Cline, T. (2000) Multilingualism and dyslexia: Challenges for research and practice. *Dyslexia: An international journal of research and practice*, 6(1), 3-12.
 Cline, T. and Frederickson, N. (1999) Identification and assessment of dyslexia in bi/multilingual children. *International Journal of Bilingual Education and Bilingualism*, 2(2), 81-93.
 Cline, T. and Shamsi, T. (2000) Language needs or special needs? The assessment of learning difficulties in literacy among children learning English as an additional language (Research Report RR184). London: Department for Education and Employment.
 Durkin, C. (2000) Dyslexia and bilingual children – Does recent research assist identification? *Dyslexia: An international journal of research and practice*, 6(4), 248-267.
 Gunderson, L., D'Silva, R. and Chen, L. (2011) Second language reading disability: International themes. In McGill-Franzen, A. and Allington, R.L. (Eds) *Handbook of Reading Disability Research*. Oxford and New York: Routledge, pp.13-24.
 Peer, L. and Reid, G. (Eds) (2016) *Multilingualism, literacy and dyslexia: A challenge for educators*. London: Routledge.
 Tsagari, D. and Spanoudis, G. (Eds) (2013) *Assessing L2 students with learning and other disabilities*. Newcastle upon Tyne: Cambridge Scholars Publishing.

3 Understanding results

3.1 Types of report

Exact creates a report for each student, which contains various sections. These are outlined below.

3.1.1 Results profile

This shows the standard scores for the six different measures (word recognition, reading comprehension accuracy, reading comprehension speed, spelling, typing speed and handwriting speed) in both 'graphical' and 'tabular' form. The key data needed for JCQ Form 8 may be extracted from these. The summary table of results also shows percentile scores for each test. Note that on the chart the average score range (standard score 85–115) is shaded grey. To aid speedy identification of areas of difficulty, the bars on the chart are coloured blue if the standard score is 85 or above (i.e. within the normal range or better), and yellow if below 85 (i.e. below the normal range, indicating that the result is a matter of concern). An example is shown in Figure 1. For explanations of the meaning of 'standard scores' and 'percentile scores' see Sections 3.2.1 and 3.2.3, respectively.

3.1.2 Cautionary warning regarding dubious reading speeds

The program checks whether the student has devoted a reasonable amount of time to the reading comprehension passages. If a student has completed the reading comprehension test in **less than eight minutes** the results should be regarded as '**doubtful**', i.e. it is unlikely that proper consideration has been given to the answers, and hence the scores will be unreliable and should not (on their own) be used as meaningful evidence for exam access arrangements. If a student completes the reading comprehension test in **less than five minutes**, the results should be regarded as '**impossible**', i.e. the student has answered the comprehension passages so quickly that it is impossible for them to have given proper consideration to the answers, and hence the scores are not safe to be used as evidence for any purpose.

When the program detects doubtful or impossible performance, a warning is given in red underneath the summary table, and the bars relating to that performance are shown in coloured hatching rather than solid block colour. Since this outcome necessarily places limitations on the use that can be made of the results of the reading comprehension test, assessors may wish to repeat this test having provided appropriate guidance to the student regarding how the test should properly be attempted (see Section 2.2 for further advice on this matter). When re-testing, the alternate form of the test (A or B, as appropriate) should be employed (see Section 2.9 for guidance on this).

3.1.3 Assessor's comments

There is a space at the foot of the Results profile for assessor's comments, which can be typed into Testwise. As a rough guide, about 1,250 characters may be included in the comment. The report will not check the length of the text entered, so it may overflow the page if too many words are entered. Alternatively, comments may be typed separately and pasted in, or written directly on to the Exact printout.

3.1.4 Results breakdown

This gives a complete breakdown of all the test scores in several tables, including comparison of ability to read and spell regular and irregular words, and the complete passage as typed to dictation by the student. An example is shown in Figure 2. Results on this page are shown in the following principal formats: standard scores (for explanation see Section 3.2.1), confidence intervals (for explanation see Section 3.2.2), percentile scores (for explanation see Section 3.2.3) and age equivalents (for explanation see Section 3.2.4). In addition, this page includes raw scores (or, in the case of word recognition, transformed scores – see Sections 1.4 and 2.1 for an explanation regarding this) and (where appropriate) times.

When the program detects doubtful or impossible performance on the reading comprehension test a warning is given in red underneath the results breakdown for that test (for further information see Section 3.1.2).

3.1.5 Checking the scores from the dictation tests

The raw scores for the dictation tests (i.e. the number of words typed and handwritten) are estimated by the computer based on the typed text saved and the number of phrases listened to by the student for the handwriting task. In about 95% of cases these figures are sufficiently accurate to be safely used in the report. However, in a few cases where the student has not followed the instructions properly the computer's estimates can be significantly different to the true figures.

Administrators should therefore carry out a visual inspection of the number of words typed and handwritten and, if a discrepancy is suspected, the administrator can manually count the number of words and enter the figures into Testwise. The manually entered figures will then replace the computer's estimated figures in the report.

This page will also show the number and percentages of spelling errors in handwriting and typing if the raw data for these have been entered. The computer does not count or estimate the number of spelling errors made in the dictation tasks and hence, if this information is required on the report, the administrator must manually enter the relevant data. The procedure for this is shown on the Testwise help site. The program will then calculate the percentages of spelling errors and display these on the report. If the relevant data are not manually entered, the number and percentages of spelling errors in handwriting and typing will be shown as zeros on the report.

3.1.6 Itemised responses

This gives the student's responses for all items in reading comprehension and spelling. This information can be useful for diagnostic purposes. An example is shown in Figure 3. Note that as the spelling test is adaptive, not all items are administered; skipped items are shown as a dash but are credited to the score as if passed correctly.

3.2 Types of scores

All raw scores on Exact are saved automatically to a single data file on completion of each test. The data saved also include the date and time the test was completed, as well as the registered details of the student. If a test has been abandoned before completion, then no results will be saved for that test.

The program then refers to the standardised norms in order to convert raw scores to the following three types of score:

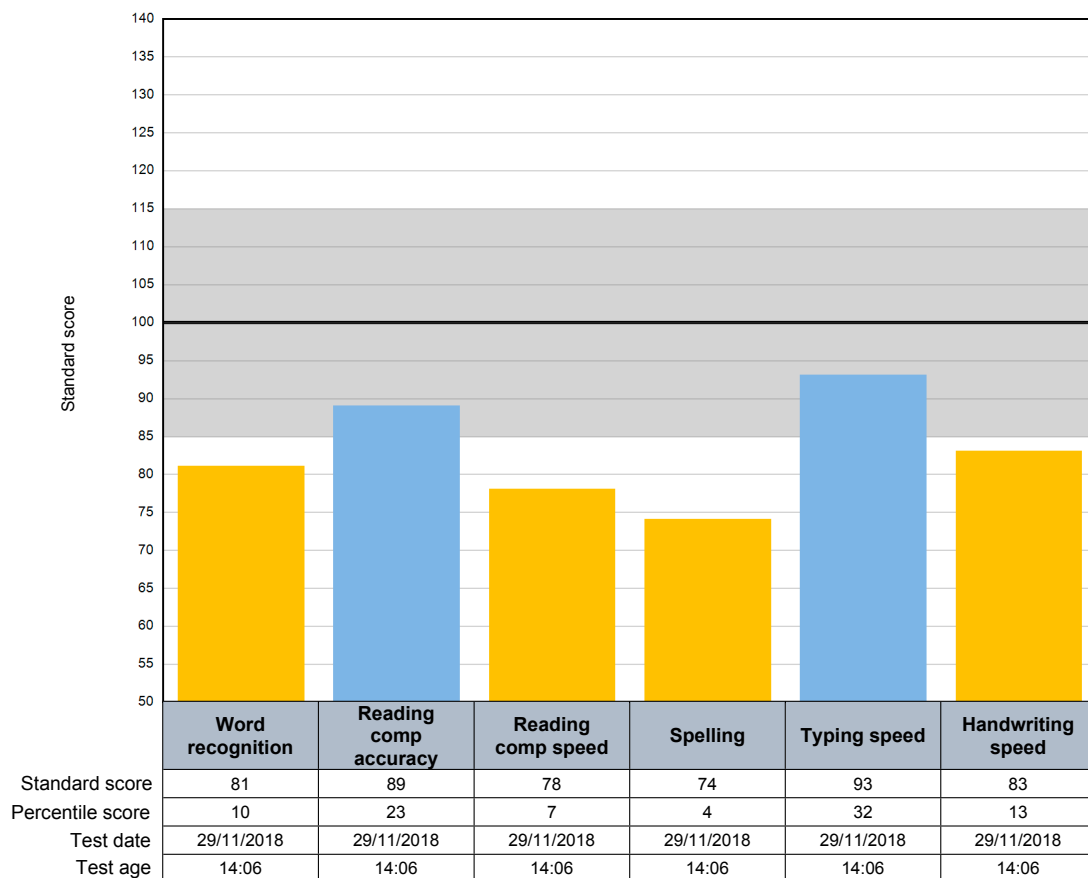
- Standard scores (and confidence intervals)
- Percentile scores
- Age equivalents

The first of these is shown in graphical (bar graph) format as well as numerical format, while the remaining two are shown only in numerical format. These different types of score formats are explained in the following sections.

Name: Student John	
Date of birth: 08/05/2011	Form: A
Date of test: 29/11/2025	Group: Class1

Results profile

Standard score of 85 or higher Standard score of 84 or lower



Assessor's comments John is a bright boy who has a diagnosis of dyslexia. His reading, writing and spelling skills have been noted by teachers to be below levels expected for his age and general ability, and this is confirmed by the tests, with all literacy measures except reading comprehension and typing speed below standard score 85. John tries to use his intelligence to compensate for his poor reading skills, with the result that his reading comprehension score is within the average range. However, his reading speed is very slow and he has difficulty decoding new or unusual words, especially under conditions of time pressure as in examinations, which often causes him to misunderstand the questions. His spelling and writing are particularly poor, with about 13% misspelled or undecipherable words. Although his handwriting speed is below average, his typing speed is in the average range for his age. It is recommended that John be permitted 25% extra time in examinations and also to use a word processor.

Figure 1. Example of Results profile

Name: Student John	
Date of birth: 08/05/2011	Form: A
Date of test: 29/11/2025	Group: Class1

Results breakdown

Word recognition

Test date: 29/11/2018 **Age at test:** 14:06

Measure	Regular words	Irregular words	Overall words
Standard score	88	75	81
Confidence interval	84 - 92	70 - 80	75 - 87
Percentile score	21	5	10
Transformed score	12	9	21
Age equivalent	12:0 - 12:11	9:4 - 10:3	11:4 - 11:7

Total words correct: 39

Total time (min:sec): 03:11

Reading comprehension

Test date: 29/11/2018 **Age at test:** 14:06

Measure	Accuracy	Speed
Standard score	89	78
Confidence interval	79 - 95	68 - 88
Percentile score	23	7
Raw score	24	50 wpm
Age equivalent	12:8 - 13:3	9:7 - 9:11

Passage number	1	2	3	4	5	Total
Questions	5	10	10	10	10	45
Attempts	5	10	10	10	3	42
Correct answers	4	8	7	4	1	24
Time (min:sec)	01:00	02:37	03:03	02:25	00:56	10:00

Spelling

Test date: 29/11/2018 **Age at test:** 14:06

Measure	Regular words	Irregular words	Overall words
Standard score	67	82	74
Confidence interval	62 - 73	76 - 90	67 - 81
Percentile score	1	12	4
Raw score	14	8	22
Age equivalent	8:8 - 8:11	11:1 - 11:3	9:7 - 9:8

ADVICE: Age equivalents are approximate and should only be used where Standard scores or Percentile scores are inappropriate.

Figure 2. Example of Results breakdown

Exact Individual student report



Name: Student John	
Date of birth: 08/05/2011	Form: A
Date of test: 29/11/2025	Group: Class1

Written dictation**Test date:** 29/11/2018 **Age at test:** 14:06

*Words dictated / actual count	128
Speed	18 wpm
Standard score	83
Confidence interval	75 - 91
Percentile score	13
Age equivalent	12:0 - 12:5
*Spelling errors (count)	17
Spelling errors (per cent)	13.00

* Actual figures can be entered manually by the assessor – see the [Testwise help site](#) for details

Typed dictation**Test date:** 29/11/2018 **Age at test:** 14:06

*Words dictated / actual count	147
Speed	21 wpm
Standard score	93
Confidence interval	86 - 100
Percentile score	32
Age equivalent	13:0 - 13:5
*Spelling errors (count)	21
Spelling errors (percent)	14.00

* Actual figures can be entered manually by the assessor – see the [Testwise help site](#) for details

ADVICE: Age equivalents are approximate and should only be used where Standard scores or Percentile scores are inappropriate

Typed passage

one day last year we saw a lot of bees going in and out of some hold hifes behind our garige. My frend once sat on a bee and got stong on the bottem so i was a bit scared of them but I know a lady who keeps bees and i asked her if she will take them away. However she thortlernt to keep them myself. She told me what i needed and what cloves to buy the was a speshal jacket, which included a hat and vayal, and some thick lever gloves, then she sowed how the bees beld their come on trans in the hive and she tote me hoe to handel them. Then we lifted off the roof of the hife and looked at the fraims where the quean was laying her eggs. When we spotted her, we marked her with a dab

Name: Student John	
Date of birth: 08/05/2011	Form: A
Date of test: 29/11/2025	Group: Class 1

Itemised responses

Reading comprehension accuracy

Passage 1

Q	Correct	Response	Score
1	No	No	1
2	Yes	Yes	1
3	Yes	Yes	1
4	No	No	1
5	Impossible to say	Yes	0

Passage 2

Q	Correct	Response	Score
1	No	No	1
2	Impossible to say	Impossible to say	1
3	No	No	1
4	Impossible to say	Impossible to say	1
5	No	No	1
6	Jay	Jay	1
7	Impossible to say	Impossible to say	1
8	No	Impossible to say	0
9	Yes	No	0
10	Yes	Yes	1

Passage 3

Q	Correct	Response	Score
1	No	No	1
2	Yes	Impossible to say	0
3	Impossible to say	Impossible to say	1
4	Impossible to say	Impossible to say	1
5	Yes	No	0
6	Impossible to say	Impossible to say	1
7	No	No	1
8	Main meals	Main meals	1
9	Yes	No	0
10	No	No	1

Passage 4

Q	Correct	Response	Score
1	No	Yes	0
2	Yes	Yes	1
3	No	No	1
4	Impossible to say	Impossible to say	1
5	Yes	Yes	1
6	No	Impossible to say	0
7	Yes	No	0
8	Impossible to say	Yes	0
9	No	Yes	0
10	Impossible to say	Yes	0

Passage 5

Q	Correct	Response	Score
1	No	Yes	0
2	Yes	No	0
3	Impossible to say	Impossible to say	1
4	Impossible to say	Not attempted	0
5	Yes	Not attempted	0
6	Previous events	Not attempted	0
7	No	Not attempted	0
8	No	Not attempted	0
9	Impossible to say	Not attempted	0
10	Yes	Not attempted	0

Figure 3. Example of itemised responses

Name: Student John	
Date of birth: 08/05/2011	Form: A
Date of test: 29/11/2025	Group: Class1

Spelling accuracy (test is adaptive)

Q.	Correct	Spelt as
P1	bed	bed
P2	nose	nose
P3	juice	joos
P4	increase	---
P5	crumb	---
6	foot	---
7	name	---
8	top	---
9	day	---
10	man	---
11	pan	---
12	egg	---
13	ice	---
14	sea	---
15	kick	---
16	girl	---
17	goal	---
18	web	web
19	knife	knife
20	pair	pear
21	nature	nature
22	diagram	diergam
23	infection	infeckion
24	engine	engine
25	ghost	ghost

Q.	Correct	Spelt as
26	captain	captain
27	acrobat	akrobat
28	festival	festeval
29	height	height
30	physics	pysics
31	theatre	theatre
32	sphere	sfere
33	anchor	---
34	referee	---
35	orchestra	---
36	aquarium	---
37	mysterious	---
38	mosquito	---
39	binoculars	---
40	observatory	---
41	literature	---
42	photosynthesis	---
43	environment	---
44	architecture	---
45	circumference	---
46	necessary	---
47	rhythm	---
48	hygienic	---
49	catastrophe	---

3.2.1 Standard scores

Standard scores are provided in 1-month age bands from 13:0 to 18:11. Standard scores have a mean (average) of 100 and a standard deviation of 15.¹⁵ They are distributed in a normal (bell-shaped) curve as shown in Figure 5. Distribution of Exact scores on a normal curve. Approximately two-thirds of the population will have scores that fall between plus or minus one standard deviation of the mean (i.e. score range 85–115, which is the area shaded blue on the graph in Figure 5. Distribution of Exact scores on a normal curve (for explanation see Section 3.2.1). In some scoring systems the range 85–115 is regarded as the ‘average range’, while other systems treat 90–110 as the ‘average range’; in the latter case, 50% of the population will fall into the average band. The more extreme the score the fewer individuals are found in that category, so that only about 2% of the population have very low scores (less than 70) and about 2% have very high scores (130+). This distribution of scores is a characteristic of all human attributes (height, weight, strength, sociability, etc.), i.e. most people tend to cluster around a central point, and as one approaches the extremes (known as the ‘tails’ of the distribution) fewer people are found.

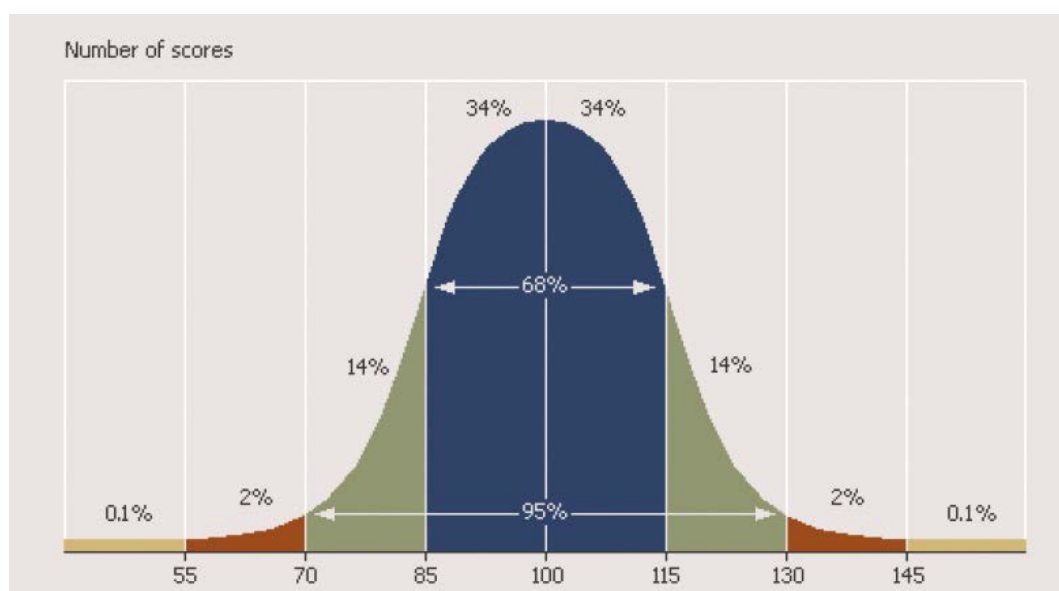


Figure 5. Distribution of Exact scores on a normal curve (for explanation see Section 3.2.1)

3.2.2 Confidence intervals

When reporting a standard score, it is good practice also to report the *confidence interval* attached to that score. The reason for this is that all psychological and educational test scores give only *estimates* of ability, based on a sample of behaviour at a given point in time. If you were to assess a student on several occasions you would not expect him or her to obtain exactly the same score each time – there would be a spread of scores, and somewhere within that spread we would expect the (hypothetical) *true score* to lie. The amount of spread or variation of actual scores obtained by an individual is dependent on the *reliability* of the test. The confidence interval is the zone around the standard score in which we are reasonably confident the *true score* lies. Different confidence intervals may be set; for Exact we have set a confidence level of 90%, which means that there is a 90% probability that the true standard score lies within the stated

¹⁵ The standard deviation is a measure of the variation of scores in a distribution.

confidence interval. Put another way, if the student was retested 100 times, on 90 out of 100 occasions the score would lie within the stated confidence interval.

Confidence intervals are calculated on the basis of the 'Standard Error of Measurement' (SEM) of a test which, in turn, is determined by the reliability of the test and the standard deviation of test scores (see Section 1.6). Confidence intervals are shown in the Results breakdown section of the Exact report.

3.2.3 Percentile scores

Percentile scores are provided in 3-month age bands from 13:0 to 18:11. Percentile scores (sometimes referred to as 'centile' scores) represent the student's performance in comparison with the population norms in percentile units, which range (roughly) from 1 to 99. A percentile score of 63, for example, means that the student's score lay at the point where 63% of the population scored less, and 37% scored more. A percentile score of 50 indicates that the student's score lay exactly on the median (middle point) of the distribution, with half the age group scoring higher and half lower. As will be obvious from Figure 5, percentile scores have a strict relationship with standard scores, as shown in Table 5. Relationship between standard scores and percentile scores.

Table 5. Relationship between standard scores and percentile scores

Standard score	70	80	85	90	100	110	115	120	130
Percentile score	2	9	16	25	50	75	84	91	98

3.2.4 Age equivalents

Age equivalents are provided for the age range 6:0 to 18:11 (over this age, age equivalents become meaningless). Age equivalents may be defined as the average chronological age of students who would be expected to achieve a given raw score on the test. The most common type of age equivalent in educational testing is the 'reading age'. For example, to say that a student has a reading age of 14 means that they read like an average 14-year-old, regardless of their chronological age. Note that age equivalents, by their very nature, are not as precise as standard scores or percentile scores; age equivalents are approximations and hence are often given in bands. Age equivalents should be used with caution and only in cases where standard scores or percentile scores would be inappropriate or unhelpful. It is embarrassing and demotivating for a teenager or adult to be told (for example) that they have a spelling age of a 7-year-old! However, some teachers working in special education prefer to use age equivalents rather than percentile scores, because age equivalents enable them to conceptualise the ability level of the student they are teaching, and so pitch the work at the correct level.

4 Use of Exact when applying for access arrangements

4.1 JCQ regulations

In England, Wales and Northern Ireland, applications for access arrangements for GCSE and GCE examinations are now made online using the Access Arrangements Online (AAO) tool. Centres are required to record assessment information on the Joint Council for Qualifications' Form 8 (available at www.jcq.org.uk).

4.2 Guidelines on using Exact

Guidelines to clarify how Exact may be used when assessing students for examination access arrangements can be found by following this link: <https://support.gl-assessment.co.uk/knowledge-base/assessments/exact-support/downloads/downloads>

5 Appendices

5.1 Typed dictation texts for FORM A

FORM A – Typed text

Bee-keeping

One day last year we saw a lot of bees going in and out of some old boxes behind our garage.

My friend once sat on a bee and got stung on the bottom, so I was a bit scared of them. But I know a lady who keeps bees, and I asked her if she would take them away. However she thought that I should learn to keep them myself.

She told me what I needed and what clothes to buy. There was a special jacket, which included a hat and veil, and some thick leather gloves. Then she showed how the bees build their comb on frames in the hive, and she taught me how to handle them.

We used a smoker to puff out smoke which calms them down. Then we lifted off the roof of the hive and looked at the frames where the queen was laying her eggs. When we spotted her, we marked her with a dab of white marker ink on her body. We made certain that she was in the bottom box of frames and put a wire mesh called a queen excluder, on top of that box. This keeps her from laying eggs in the boxes above, where the other bees store honey.

[212 words]

FORM A – Handwritten text

My Dog, Bill

My dog Bill is very old, but he still likes to play football. We go out into the garden, and Bill tries to get the ball away from me and run off with it.

On Sunday he was 10, so we wanted to give him a birthday treat. My brother Tom said that we should leave him out in the front garden, where he could bite the postman. However, my mum said that it was a really stupid suggestion, as Bill was in enough trouble already from chasing the neighbour's cat.

In the end we decided to take him to the park which he always enjoys because he can chase the pigeons and play with other dogs.

On that day, there was this really posh poodle dressed up in a tartan jacket. You could see that Bill thought she was terrific. He couldn't take his eyes off her and he kept bouncing around in front of her as if he was a puppy.

She was having absolutely nothing to do with him and walked away with her nose in the air. Poor Bill. To make up for his disappointment, we bought him one of those artificial bones that dogs love to chew.

[201 words]

5.2 Typed dictation texts for FORM B

FORM B – Typed text

Biking

My name is Ben and I live in the country. If I want to see my mate Joe, I have to ride my bike, because he lives 2 miles away on a farm.

They have a quad bike there for looking after the sheep, and we are allowed to ride it all round the place. We have made a track over a hill, through a wood and down the side of a field to the farmyard. At the end, there is this really sharp bend between a combine harvester and a derelict tractor, where it is very bumpy and I usually fall off.

Yesterday I completed the course in 4 minutes 20 seconds, which is a record. I almost fell off at the last corner, but I managed to hold on to the handlebars, run beside the bike for a few paces and jump on again. It was like you see the bobsleigh teams do on TV.

Normally if you let go of the throttle, the bike just stops. However last week when Joe came off, it jammed open and the bike went at full speed into some hay bales. Luckily it wasn't damaged and we didn't let on to his dad what had happened.

[206 words]

FORM B – Handwritten text

The Lock

Sam was 14 and he knew it all.

When the family went on the river in a long boat, Sam told them what to do, even Mum and Dad. He told Rose and Jamie not to fall in the water, and Mum how to push off the boat. He even told Dad where to steer.

When they got to the first lock, Sam knew how to work it. They tied up just below it, while Sam and Mum opened the gates and Dad took the boat into the lock. Then they closed the sluices on the lower gates and opened the upper ones.

Sam showed Rose how to wind up the sluices with the big handle on the top of each gate. He stepped back to watch her and lost his balance on the narrow footway, falling backwards into the lock with a tremendous splash. Dad fished him out with the boat hook. He was filthy, wet and furious, but everyone else just laughed including the lock keeper, and people from two cruisers.

[190 words]