Aptis Technical Update 2015-16

Developed by the Assessment Research Group (ARG)

This annual summary presents key statistics for the British Council's English language assessment test, Aptis, from the financial year 2015-16 (April-March).

The report gives an overview of the trends in test functioning and usage for the main Aptis variants operationally live during this period: Aptis (computer-based and browser version), Aptis for Teachers, and Aptis for Teens.

Aptis tests delivered

During the period April 2015 – March 2016, Aptis computer-based tests were delivered in more than 80 countries.

This included a number of different "variants" of the test, i.e. versions of the test that have been developed for different audiences. This process is operationalised in terms of levels of localisation, explained in greater detail in the technical manual (O'Sullivan and Dunlea, 2015). In essence, whilst the fundamental role of the test is to evaluate test-takers' capability in the English language, the test instruments can be 'adapted for use in particular contexts with particular populations to allow for particular decisions to be made' (ibid., p.7). This speaks to the flexibility of the test for the end users.

Proportion of tests delivered by (major) variants

![British Council Aptis logo - Forward thinking English testing]
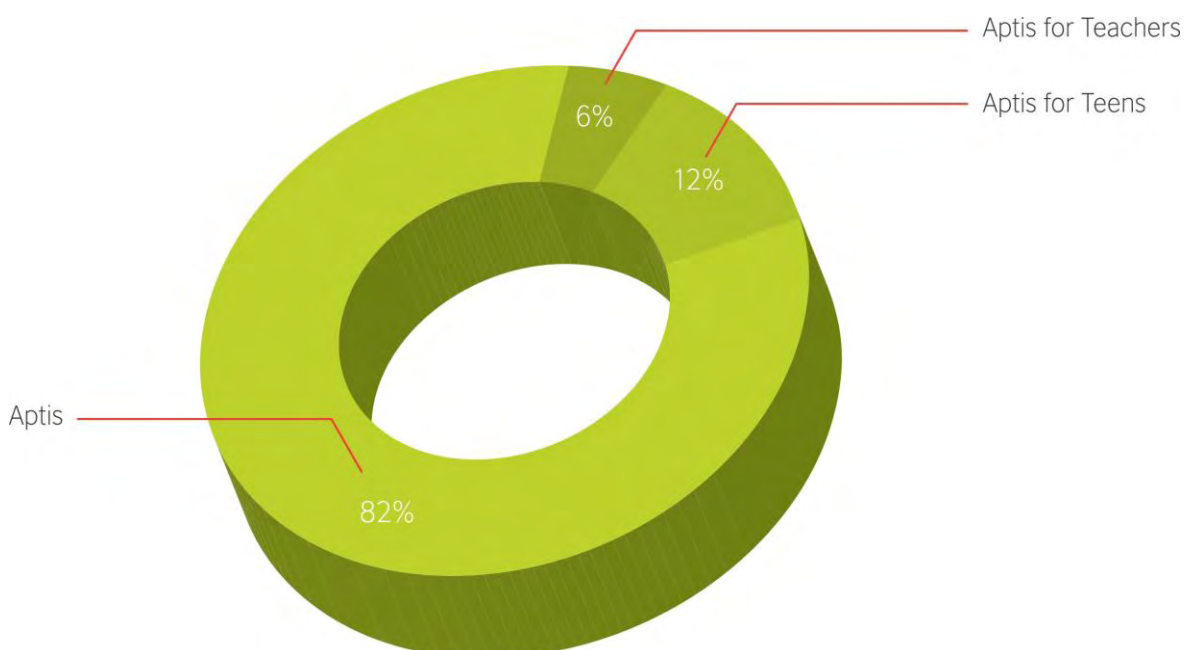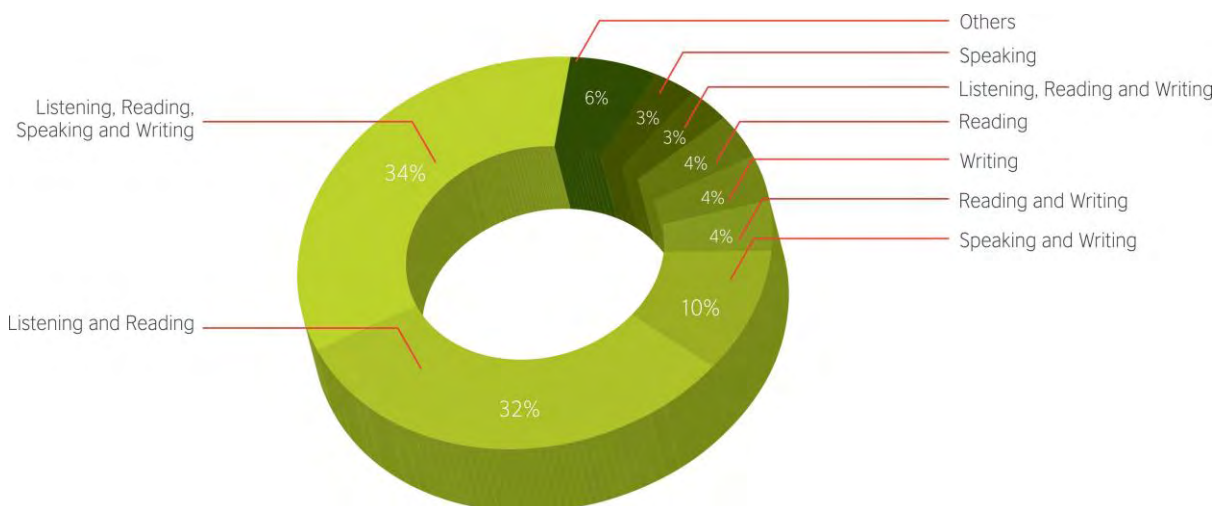
Aptis test packages

Test packages include different combinations of the four skills: Reading, Listening, Writing and Speaking. Each package will always include the Core (Grammar and Vocabulary) component. This optionality allows organisations to choose a package that focuses resources on those skills most relevant to their needs.

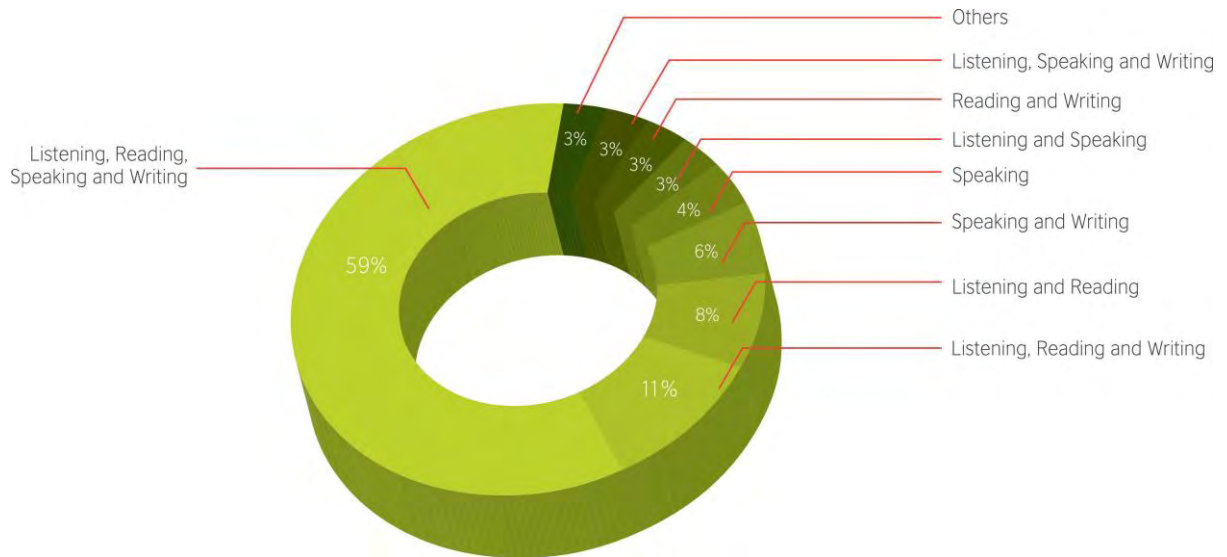Pie charts show the breakdown of packages delivered across the three major variants.

> The four-skill package (Listening, Reading, Writing and Speaking) is most popular for the audience-specific variants (Teens and Teachers).

Aptis (computer version) – test packages

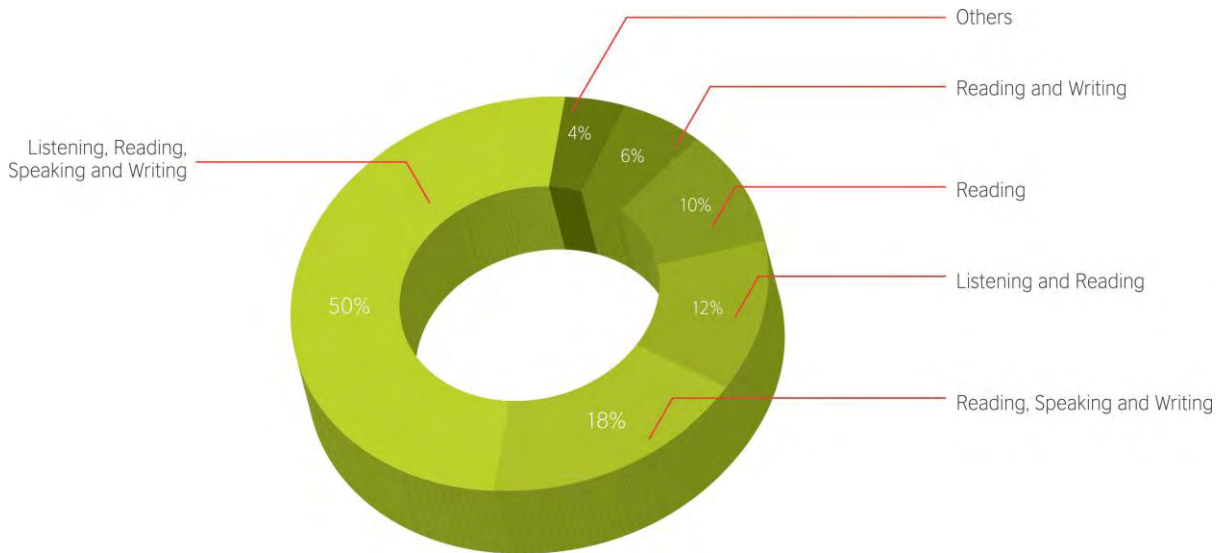Pie chart legend and values:
- Listening, Reading, Speaking and Writing: 34%
- Listening and Reading: 32%
- Speaking and Writing: 10%
- Reading and Writing: 4%
- Writing: 4%
- Reading: 4%
- Listening, Reading and Writing: 3%
- Speaking: 3%
- Others: 6%

Aptis (browser version) – test packages

- Others — 3%
- Listening, Speaking and Writing — 3%
- Reading and Writing — 3%
- Listening and Speaking — 3%
- Speaking — 4%
- Speaking and Writing — 6%
- Listening and Reading — 8%
- Listening, Reading and Writing — 11%
- Listening, Reading, Speaking and Writing — 59%

Aptis for Teachers – test packages

- Others — 8%
- Reading and Writing — 3%
- Listening, Reading and Writing — 4%
- Reading — 15%
- Listening, Reading, Speaking and Writing — 70%

![British Council Aptis logo]
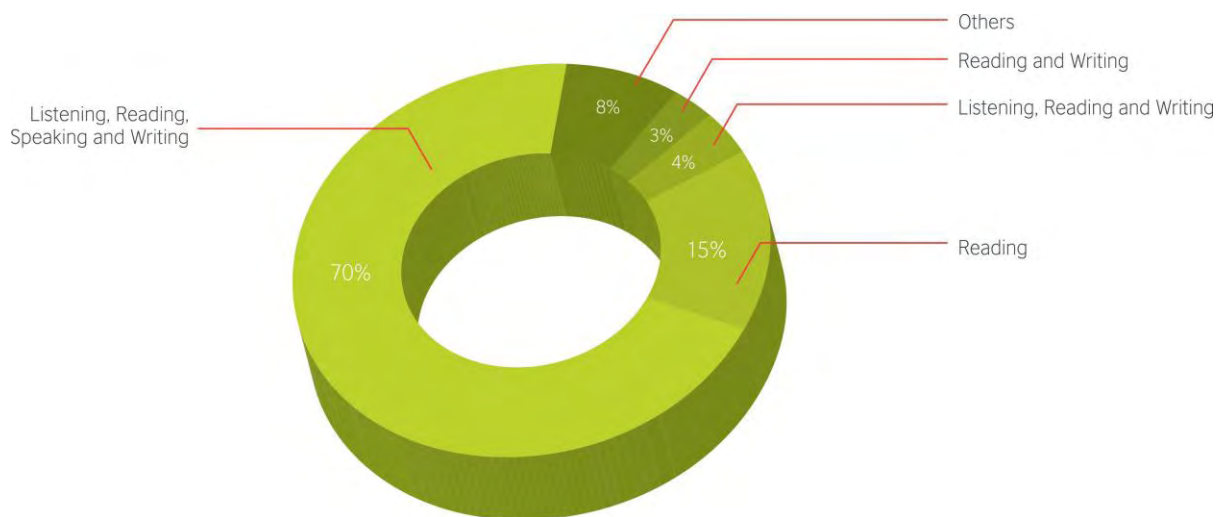
Aptis for Teens – test packages

- Others — 4%
- Reading and Writing — 6%
- Reading — 10%
- Listening and Reading — 12%
- Reading, Speaking and Writing — 18%
- Listening, Reading, Speaking and Writing — 50%

Aptis test performance statistics

Detailed analysis was carried out by the Assessment Research Group (ARG) on the reliability of each skill-area component of each test variant.

The most appropriate means of calculating reliability depend on the format in which task and item responses are recorded (Cronbach's Alpha for Core and Listening components; KR21 for Reading).

Reliability statistics give an indication of what is known as the "internal consistency" of a test, essentially this refers to whether all the items are working together to assess the same underlying ability.

The most appropriate means of calculating reliability depend upon the format in which task and item responses are recorded. In all cases, the reported scale is between 0-1, with 1 representing a higher degree of reliability. A figure greater than 0.8 is generally considered to indicate that a set of test items are functioning as required. For the productive skills (Speaking and Writing) a measure of "inter-rater" reliability is given. This indicates how consistently the examiners are marking. Again scores closer to 1 are desirable.

In addition to reliability statistics, SEM (Standard Error of Measurement) is reported for each component. SEM figures indicate the certainty with which we can be sure that score achieved represents the test-taker's level of ability in that particular area by indicating the

range in which a given test-taker's score will fall in a retest situation. There is an acknowledged trade-off between this statistic and the inclusion of a variety of item types. All tests in the Aptis suite are on a scale out of 50, and the SEM figures show the range either side of which we can be confident that the candidates' score would fall were they to take the test again.

Table 1: Reliability and SEM for Core component (Grammar and Vocabulary)

Variant name	Aptis	Browser	Aptis for Teachers	Aptis for Teens
Reliability*	0.92	-	0.91	0.94
SEM	2.99	-	3.36	2.91

Cronbach's Alpha

Table 2: Reliability and SEM for Reading

Variant name	Aptis	Browser	Aptis for Teachers	Aptis for Teens
Reliability*	0.88	-	0.84	0.87
SEM	4.19	-	4.41	4.35

KR21

Table 3: Reliability and SEM for Listening

Variant name	Aptis	Browser	Aptis for Teachers	Aptis for Teens
Reliability*	0.849	-	0.87	0.83
SEM	4.04	-	3.84	4.10

Cronbach's Alpha

Table 4: Reliability and SEM for Writing

Variant name	Aptis
Reliability*	0.92
SEM	3.69

Inter-rater reliability; see O'Sullivan and Dunlea (2015, p.22) for calculation details.

Table 5: Reliability and SEM for Speaking

Variant name	Aptis
Reliability*	0.90
SEM	4.44

Inter-rater reliability; see O'Sullivan and Dunlea (2015, p.22) for calculation details.

Overall, the information in Tables 1–5 shows that the test is performing within acceptable tolerances for tests of this nature and scale across all components.

All Cronbach's Alpha and KR21 figures reported are above a desirable 0.8 for a scale of 0-50. In addition, inter-rater reliability statistics give a positive indication of rater performance for Aptis.

Overview

Test performance figures reported in this document are comfortably within the target ranges, whilst delivery statistics show that the Aptis test has been taken up on a global scale, with high volume countries in Europe, Asia, Africa, the Middle East and South and Central America. In addition, clients are taking advantage of the flexibility offered by Aptis, as seen by the delivery of Teens and Teachers variants, as well as the range of packages that have been supplied in 2015-16. Overall, this valuable information about the reach and functioning of the test sets good ground for the upcoming 5th anniversary of Aptis in August 2017.