Factor structure and four-skill profiles of the Aptis test

This is a summary of a report by Yo In'nami and Rie Koizumi as part of the ARAGs Research Online Series. See the report at: www.britishcouncil.org/exam/aptis/research/publications/

WHAT WE LOOKED AT:

We investigated the factor structure and skill profiles of examinees of the Aptis test. The factor structure of a test exhibits the abilities assessed in the particular test. The Aptis test aims to assess the four language skills – listening, reading, speaking and writing. Does the Aptis test data show that the test assesses these four skills (statistically modelled as four factors)? How are the four skills (factors) related to each other? To what degree is the skill (factor) structure shown in the data in line with the structure that the Aptis team intends to put into operation in their test? Further, is the structure the same as predicted by theory, i.e., is the structure a multicomponential structure of language ability, which means that language ability consists of two or more related factors? Are the results same across countries? These questions are addressed in Study 1.

Two research questions were examined in Study 1:

1. What is the factor structure of the Aptis test when all countries are analysed together, as well as individually?
2. To what extent does measurement invariance hold across countries? (i.e., to what extent can we observe the same results in terms of the factor structure across countries?)

In Study 2, we examined the Aptis examinees' skill profiles. Previous studies have shown that test results reveal either flat or uneven profiles (e.g., skill levels are similar or different). By investigating these profiles, we can identify the strengths and weaknesses of examinees and classify them into different learner groups according to the profile type.

One research question was examined in Study 2:

3. Are there different language profiles across learner groups? If so, what different language profiles are present?

HOW WE DID IT:

We analysed international comparison data collected by the British Council which included test scores for four skills (i.e., listening, reading, speaking and writing) of 1,270 examinees in seven countries: Bangladesh, Chile, Indonesia, Mexico, Poland, Spain and Sri Lanka.

In Study 1, we used confirmatory factor analyses and multiple-sample analyses across countries. In Study 2, we used latent profile analyses.

WHAT WE FOUND:

The major findings and implications are as follows:

Research Question 1: What is the factor structure of the Aptis test when all countries are analysed together, as well as individually?

When all seven countries were analysed together, the factor structure that best explained the data was a bi-factor model, where a single, general second language (L2) proficiency factor, as well as the four skills, influenced the test scores directly. This result held true for Bangladesh. In contrast, the other six countries (Chile, Indonesia, Mexico, Poland, Spain, and Sri Lanka) had a factor structure that was best explained by a higher-order model, where a single, general L2 proficiency factor influenced the four skills, which in turn influenced the test scores. These two models are similar in that the test scores are explained by L2 proficiency and the four skills, both showing multicomponential factor structures, which are consistent with previous studies (e.g., Bachman & Palmer, 1982; Sawaki & Sinharay, 2018).

Research Question 2: To what extent does measurement invariance hold across countries?

Measurement invariance held when data of each country had the same factor structure. However, the strengths of relationships between test scores and factors were different (i.e., no equality constraints).

The results showed that either the higher-order or the bi-factor model was supported. We can conclude that both factor structures accord well with reporting a total score, as well as separate scores for each skill, as provided for in Aptis score reports.

Research Question 3: Are there different language profiles across learner groups? If so, what different language profiles are present?

We found five learner groups, each with a different skill profile. Three groups performed in a manner expected by the test developers (i.e., Group 2 – low group; Group 3 – intermediate group; and Group 4 – advanced group), while two did not. Of the two groups that performed in an unexpected manner, Group 1 was the low group which scored lower on easier tasks and higher on more difficult tasks of listening, reading and writing. Group 5 was an intermediate group, but demonstrated low performance across all speaking tasks.

The reasons for the unexpected profiles in Groups 1 and 5 were not related to problems with the test. Thus, the results provide diagnostic information to test-takers and developers, plus evidence of positive validity to the Aptis test users for diagnostic purposes. Further, test developers might want to consider routinely analysing examinees' skill profiles; such analysis, either simple or detailed, would reveal the examinees' skill characteristics and provide evidence of validity for diagnostic purposes.