

## Scaled scores

“Scaled scores” that have been adopted starting with the new JLPT are based on a common scale in order to make test results (scores) correspond to examinees’ Japanese-language proficiency in a more fair manner. Scaled scores enable measurement of Japanese-language proficiency based on the same consistent standard regardless of when the test is taken. The advantages and calculation method of scaled scores are summarized below.

### 【Advantages of scaled scores】

Perfectly maintaining the level of difficulty of questions among different tests is extremely difficult even if thorough analysis and discussion by experts are conducted during the question design phase. With raw scores, which are obtained simply by adding the allocated points of correctly answered questions, test scores are dependent on the difficulty of exams. When exam timing and difficulty of exam questions differ, test scores will vary for examinees with the same Japanese-language proficiency. Since examinees’ Japanese-language proficiency as expressed in raw scores is affected by the difficulty of exams, it is necessary to consider the possibility that raw scores do not appropriately reflect actual Japanese-language proficiency.

On the other hand, scaled scores adopted in the new JLPT hold great benefit for proficiency measurement methodology – “the ability to evaluate Japanese-language proficiency independently from exam difficulty and quantify it based on a standard scale.” This characteristic allows examinees with the same Japanese-language proficiency to receive the same scores regardless of when they take the test. The introduction of scaled scores also allows the “difference in scaled scores” calculated for two different tests to be interpreted separately from exam difficulty as the “difference in Japanese-language proficiency” as long as exams are for the same level and scores are for the same scoring sections.

### 【Calculation method of scaled scores】

Calculation of test scores not affected by the difficulty of exams (scaled scores) is based on a statistical test theory called Item Response Theory (IRT). This is completely different from calculation of raw test scores based on the number of correct answers.

Scaled scores are determined mathematically based on “answering patterns” of how examinees answer particular questions (correctly or incorrectly). For example, a test consisting of 10 questions (items) has a maximum of 1024 answering patterns ( $2^{10}$  patterns). For the scaled-score calculation process for the new test, based on these answering patterns Japanese-language proficiency of examinees is positioned on a scale between 0 and 60 points for one scoring section (0-120 points for Language Knowledge [Vocabulary/Grammar] • Reading for N4 and N5). Because the maximum 1024 answering patterns of a test consisting of 10 questions (actual exams have more questions) are categorized into 61 groups, scaled scores sometimes become identical for different answering patterns. Therefore, scaled scores can be identical for two examinees even when the number of questions they correctly answer or their answering pattern does not

match. On the other hand, scaled scores may not be identical even though the number of correctly answered questions is the same when answering patterns are different.

For more details on the calculation method of scaled scores, please refer to the following references.

#### 【References】

Ikeda, H. (1994) *Gendai Tesuto Riron [Modern Testing Theory]*. Tokyo: Asakura Publishing.

Toyoda, H. (2002) *Komoku Hanno Riron Nyumon-hen – Tesuto to Sokutei no Kagaku [Item Response Theory Introduction – Test and Measurement Science]*, Tokyo: Asakura Publishing.

Watanabe, N. & Noguchi, H. (eds.) (1999) *Soshiki Shinrigaku Sokuteiron – Komoku Hanno Riron no Furontia [Organizational Psychology Measurement Theory – Frontier of Item Response Theory]*, Tokyo: Hakuto-Shobo Publishing Company.

#### **Setting passing points**

The passing standard for the new test is based on statistical analysis and designed to match that of the old test. Examinees able to pass Levels 1, 2, 3 and 4 in the old test are judged to have the Japanese-language proficiency to be able to pass N1, N2, N4 and N5 in the new test. As for N3, there was no corresponding level in the old test, and passing scores for N3 are designed to fall between the passing standards of the old test's Levels 2 and 3, based on statistical analysis of Japanese-language proficiency levels for passing these levels.