

Свирипчук И.А.

Национальный технический университет Украины «Киевский политехнический институт им. Игоря Сикорского»

REQUIREMENTS OF LANGUAGE TESTS

Language testing is central to language teaching. It provides goals for language teaching, and it monitors, for teachers and learners, success in reaching those goals. Its influence on teaching is strong – and is usually felt to be wholly negative. It provides a methodology for experiment and investigation in both, language teaching and language learning.

It is normal for teaching to be directed towards assessment, as examination or test. References used in language testing discussion often refer to psychological work. It is therefore of interest to consider to what extent language tests are a sub-category of psychological tests. Many of the principles and procedures involved in language test construction derive from psychological testing. Indeed there is a view which says that a language test is psychological test with a special content and that therefore constructors of language tests need to learn about testing rather than about language. Anstey suggests four stages of the test. They are:

1. Plan the content and general layout of the test, decide on the type of test item, the length and time limit for the test in its final form, the instructions to be given and the methods of scoring. This is the planning stage.
2. Devise at least three times as many items as will eventually be needed (more if two parallel versions are likely to be required) and try the rough draft on a small group of interested people in order to obtain introspections on the general impact of the tests, and to identify items which are palpably unsatisfactory. This is pre-pilot stage.
3. A try-out, preferably two or more try-outs of successive drafts, on a large sample of the same kind of people on whom the test is to be used, in order to

check the test administration and provide material for thorough item analysis and revision of the draft test. This may be called the pilot stage.

4. Try-out of the test in its final form in order to obtain evidence as to its practical usefulness or validity and to obtain “norms”, that is means of assessing the significance of scores in the test. This is the final validation stage.

In order to show the extent of a test’s reliability the most obvious way is to construct an exactly similar versions of the test, to pilot that on the same sample as the first version and then to compare results. A highly reliable test will have complete agreement (or very nearly so) between the two versions. Of course there are problems – such as ensuring that the sample under test is behaving in the same way at each test administration – but approximations can be made. The greater problem is to be sure that one test is in fact exactly similar to another and while this can be statistically achieved (through comparable item indices) there seems to be no safe way in which we can be sure from a language point of view that one test item is equivalent to another. A second alternative is to use exactly the same test and repeat it with the same objects. The assumption here is that, while all subjects will do better because of practice, they will all do so equally well. A third alternative is to give the test only once but to behave as if it was in fact two tests, two versions, which happened to have been administered together. The test can then be split in half, the two halves correlated and then the correlation which indicates the reliability of one half can, through a straightforward boosting formula, be reinterpreted to show what reliability can be claimed for the whole test. The problem here is that there is no principled way in which a test can be split in half.

We can distinguish four test uses: achievement, proficiency, aptitude and diagnostic. The different uses can be distinguished in terms of time and content. Thus the achievement test refers back to previous learning and is concerned solely with that; achievement tests are typically used at the end of the period of learning, a school year or a whole school or college career. The content is a sample of what has been in the syllabus during the time under scrutiny.

The proficiency test is also interested in what has been learnt but in a much more vague way. Unlike the achievement test the proficiency test exhibits no control over previous learning; instead it establishes generalization on the basis of typical syllabuses leading to entry and is more directly related to what it attempts to predict, namely, performance in the language under test on some future activity.

Unlike the proficiency test the aptitude test has no content (no typical syllabuses for teaching aptitude) to draw on but like the proficiency test it is concerned to predict future achievements, though this time not in language for some other purposes (for example, practicing medicine) but in language for its own sake. An aptitude test is intended to predict future language learning success. Its design is, however, more problematic than that of either an achievement or a proficiency test since there is no body of skill or knowledge that can be sampled to produce an aptitude test. Typically, aptitude tests draw such abilities as first-language verbal knowledge, ability to codify unfamiliar phonemic features, and motivation.

Diagnostic tests are the reverse side of achievement tests in the sense that while the interest in the achievement test is in success, the interest in the diagnostic test is in failure, what has gone wrong, in order to develop remedies. There is also the widely based diagnostic test which purports to provide the detailed profile of particular areas of language learning.

ЛИТЕРАТУРА:

1. Anstey, E. (1966) Psychological tests. London: Nelson.
2. Butler, C. (1985) Statistics in Linguistics. Oxford: Basil Blackwell.
3. Davies A. (1990) Principals of Language Testing. Cambridge: Basil Blackwell.