

АНАЛИЗ СТАНДАРТИЗИРОВАННЫХ И НЕСТАНДАРТИЗИРОВАННЫХ ТЕСТОВ

Григорян Геворг

Кандидат фил. наук

Анимационный профессиональный колледж Цзянси Теллоу, Китай

gevorg.grigoryan93@gmail.com

Аннотация

Стандартные тесты являются одним из новейших и наиболее удобных методов проверки и оценки знаний учащихся. Они обеспечивают более широкий охват и более высокую объективность, позволяют обнаружить типичные ошибки и трудности студентов, дают возможность для самообучения и самооценки, в то же время позволяя совершенствовать студенческие программы и методы обучения. Тестирование является одной из самых неправильно понимаемых областей преподавания и изучения языков. Студенты и преподаватели одинаково «поёживаются», когда слышат слово «тестирование». Учащиеся видят в тестах угрозу своей компетентности, потому что боятся, что не справятся с ними. Учителя часто не любят составлять тесты и не совсем удовлетворены своими результатами, когда они это делают. Они также с подозрением относятся к стандартизированным, профессионально разработанным тестам, потому что не понимают, что эти тесты на самом деле пытаются измерить.

В последние несколько лет наблюдается растущий интерес к улучшению ситуации. Волна интереса к коммуникативному обучению языку, например, стимулировала параллельный интерес к более инновационным и чувствительным измерениям разговорной способности.

Данная статья посвящена изучению некоторых ключевых вопросов построения тестов. Чтобы правильно построить или оценить тест, полезно иметь четкое представление о том, с какой целью проводится тест. Один из способов обозначения тестов – «прогноз» – это те, которые касаются прогнозирования успеваемости учащегося, и «оценка достижений» – это те, которые оценивают текущий уровень достижений.

Ключевые слова: стандартный и нестандартный тест, оценка, тест скорости, групповой тест, индивидуальный тест, тестовые задания.

ՍՏԱՆԴԱՐՏ ԵՎ ՈՉ ՍՏԱՆԴԱՐՏ ԹԵՍՏԵՐԻ ՎԵՐԼՈՒԾՈՒԹՅՈՒՆ

Գրիգորյան Գևորգ

բան. գիտ. թեկնածու

Օլանգուի Թեղիառ անիմացիոն մասնագիտական քոլեջ, Չինաստան

gevorg.grigoryan93@gmail.com

Ամփոփում

Սույն հոդվածը նվիրված է ստանդարտ և ոչ ստանդարտ թեստերի համակարգային վերլուծությանը: Հարկ է նշել, որ ներկայումս ստանդարտ թեստերը կրթության ոլորտում դիմորդի գիտելիքների ստուգման անքակտելի տարրերից են: Առհասարակ թեստերի նկատմամբ հետաքրքրությունն ու թյուրմբոնումներն առկա են դեռևս նախորդ հազարամյակի վերջին տարիներից, երբ կրթության ոլորտում ներմուծվեցին բազմազան թեստային տարատեսակներ, որոնք հակասական կարծիքներ ձևավորեցին: Մասնագետների մի մասի կարծիքով ստանդարտ թեստերը ամբողջովին չեն արտացոլում թեստ հանձնողի գիտելիքները, նույնիսկ ավելին, դրանք սահմանափակում են դիմորդների ստեղծարարությունն ու տրամաբանական մտածողությունը, խաթարում ուսումնառության բուն գործընթացը: Մասնագետների մի այլ ստվար զանգված էլ կարծում է, որ ստանդարտ թեստերն ապահովում են գնահատման լիարժեք օբյեկտիվություն, ինչը թեստի կարևորագույն ցուցիչն է:

Սույն հոդվածում փորձ է կատարվել ներկայացնելու ստանդարտ և ոչ ստանդարտ թեստերի առանձնահատկություններն ու առկա թերությունները, ինչպես նաև դրանք կատարելագործելու ճանապարհները:

Բանալի բառեր: Ստանդարտ և ոչ ստանդարտ թեստ, գնահատում, արագության ստուգման թեստ, խմբային թեստ, անհատական թեստ, թեստային առարկաներ:

ANALYSIS OF STANDARDIZED AND NON-STANDARDIZED TESTS

Grigoryan Gevorg

Ph.D.

Jiangxi Tellhow Animation Vocational College, China

gevorg.grigoryan93@gmail.com

Standard tests are one of the newest and most suitable methods for checking and evaluating learner's knowledge. They provide broader scope and higher objectivity, they allow the typical mistakes and difficulties of the students to be discovered, they present the opportunity for self-study and self-evaluation, at the same time enabling student programs and methods of teaching to be improved.

However, testing is one of the most misunderstood areas of language teaching and language learning. Students and teachers alike cringe when they hear the word "testing". Students see tests as threat to their competence, because they are afraid that they will not perform well on them. Teachers often do not like to construct test and are not altogether satisfied with their results when they do. They are also suspicious of the standardized, professionally designed tests because they do not understand what these tests are really trying to measure.

Fortunately, in the past few years there has been a growing interest in improving the situation. The wave of interest in communicative language teaching, for example, has stimulated a parallel interest in more innovative and sensitive measures of speaking ability.

The following article is devoted to the study of some key issues in test construction. In order to properly construct or assess a test it is helpful to have some explicit notion of what the test is testing and how it might be labeled. One way of referring to tests is according to those that deal with prediction of student's performance "prognosis", and those that assess the current level of accomplishments, "evaluation of attainment".

Keywords: standard and non-standard test, assessment, power and speed test, group test, individual test, test items

Introduction. It is widely known that a test or an examination is mainly targeted to assess and measure a test-taker's knowledge, mental ability, language skills, aptitude and so forth. D. Adom, J.A. Mensah and D. A. Dake in their "Test, measurement, and evaluation: Understanding and use of the concepts in education" article mention: "One of the most commonly used assessment tools in education is to conduct tests. Beyond being considered as an instrument, tests can also be seen as standard procedures used to systematically measure a sample of behaviour by posing a set of questions [Int. J. Eval. & Res. Educ. Vol. 9, No. 1, p.110, 2020].

The novelty of this article is the determination of specific peculiarities of standard and non-standard tests.

The topicality of this research article is to evaluate the essential role of standard and non-standard tests and ascertain their significant place in the educational system.

The goal of the research is to reveal common and contrastive features of standard and non-standard tests in the education system in order to find out which works best and which will satisfy both students and teachers.

In order to achieve the above-mentioned goal, the following tasks have been set forth:

- To discuss and specify the characteristic features of standard and non-standard tests.
- To study different types of categorization of tests inherent to modern education.
- To analyze the possible transformative tendencies that will be a better match for test-takers.

The practical significance of the paper is deeply supported by the applicability of its content in education. The findings thereof can be used for lectures and seminars, as well as may serve as a theoretical basis for other research projects relevant to education.

The methods of component analysis, inquiry, comparative analysis, observation, etc., have been applied to the research to fulfill the tasks of the paper.

Tests are designed to measure the quality, ability, skill or knowledge of a sample against a given standard, which usually could be deemed as acceptable or not [Int. J. Eval. & Res. Educ. Vol. 9, No. 1, p.109-11, 2020].

The way a test is administered can vary: it can be paper-based, computer-based, oral or in a confined area that requires a test-taker to physically perform some set of skills. Undoubtedly tests vary in different aspects: style, requirements, time restriction, rigor, etc.

Every year numerous tests are used in different educational institutions, which can be classified into different types or categories. One way to classify tests is by the construct the test purports to measure. Using the above-mentioned method, tests can be organized into achievement tests, aptitude or intelligence tests, personality inventories, projective techniques, interest inventories, attitude measures and so on. In refer to this, it is worth mentioning that each of these types of tests may be further divided into subcategories. Lots of tests are classified as group or individual.

Group Tests

Individual Tests

For example, individual tests are designed for administration in one-on-one situation. Group tests are designed for a group administration. Individual tests are typically used for clinical purposes, such as making a diagnosis of a disability or disorder or determining strengths and weaknesses in a specific area of functioning (e.g. intelligence, achievement). These tests are administered to one student at a time. The test-user should have considerable training in test administration, scoring and interpretation. Group tests, on the other hand, are designed substantially as instruments for mass testing. These kinds of tests are largely pencil-paper measures suitable for administration to large or small groups of students at the same time. The majority of tests used in educational institutions are group tests.

To a great extent group tests are fairly facile to administer and score, and their use does not require much special training on the part of the examiners. Some group tests may also be computer administered and scored.

Tests may also be grouped into speed tests and power tests.

Speed Tests

Power Tests

E. Estrada in his "Speeded Tests" article mentions: "In the context of educational measurement, the term speeded test (or speed test) refers to a measuring tool composed of a list of relatively easy items, intended to be answered in a very limited time [²<https://methods.sagepub.com/reference/the-sage-encyclopedia-of-educational-research-measurement-and-evaluation/i19548.xml>].

On the whole, a speed test is designed to measure and evaluate the speed or rate of performance, rather than the acquisition or mastery of knowledge.

Normally, speed tests encompass very easy items so every test-taker knows the answers. This way, only speed of response is measured by the test. On the other hand, a power test is designed to measure the knowledge of the test-taker, regardless of his or her speed of performance.

In contrast to speed tests, power tests contain assignment-items with varying degrees of difficulty, at the same time power tests provide enough time for test-takers to attempt all items.

E. Estrada in his "Power Tests" article mentions: a power test usually refers to a measurement tool composed of several items and applied without a relevant time limit. The respondents have a very long time, or even unlimited time, to solve each of the items, so they can usually attempt all of them [³<https://methods.sagepub.com/reference/the-sage-encyclopedia-of-educational-research-measurement-and-evaluation/i16318.xml>].

Performance is based on how well a student can answer the items, instead of how fast he or she can perform. However, it should be stressed that most tests used in various educational institutions evaluate both knowledge and speed as factors on test performance. Specifically, these tests are designed to measure students' knowledge in a domain of content and a time limit is set for the students to complete the test. In other words, a student's score is influenced by both accuracy and speed of his or her answers. On such tests, although a student may have the

necessary knowledge required to answer the test items correctly, the student may not receive a high score if he or she works in a slow speed and is unable to complete many items within the time limit.

Tests are also divided into standard and non-standard types.



A standardized test is a test that is administered and scored in a consistent, or “standard”, manner. Standardized tests are designed in such a way that the questions, conditions for administering, scoring procedures, and interpretations are consistent and are administered and scored in a predetermined, standard manner

[⁴<https://www.academia.edu/38027992/StandardizedandNonStandardizedTestsSTANDARDIZEDANDNONSTANDARDIZEDTESTSNonStandardizedTests>].

Any test in which the same test is given in the same manner to all test-takers is considered to be a standardized test. Standardized tests need not be [high-stakes tests](#), time-limited tests, or [multiple-choice tests](#).

Standardized tests are perceived as being fairer than non-standardized tests. The consistency also permits more reliable comparison of outcomes across all test takers [Stewart, p.32, 1998:32].

Standardized tests, in contrast to non-standardized tests, are widely conducted in terms of scope, the level of difficulty, as well as format. It should be mentioned that generally standardized tests are carried on fixed dates, which are determined by the test developer companies, educational institutions, governing body or local authorities, which may or may not be administered by the instructor, held within the classroom, or constrained by the classroom period.

Standard way of assessment is very practical. Standardized tests are less time consuming, they are easy to administer and there is less time that is needed to administer the test rather than the other complicated assessments that require more personal time with students. There are explicit directions, so administering these tests are far easier. Machines do the grading for these tests and computers track the progress of achievement for different groups of students in different subjects. Tests are scored objectively, there is no need for teachers to have their bias or emotions to influence the score of the students. It is easy to identify problems and instigate change or reform. These problems can be easily identified on student, class, school and district level. Also, with the ease of identifying problems, there is an accountability factor for students, teachers and everyone involved in the educational system. Finally, students, regardless of socioeconomic status, linguistic or racial background or disability, should be expected to meet common standards that challenge them to acquire content and skills that are more than just minimum require.

Standardized testing affects the curriculum being taught in schools, it affects how teachers teach their students, and usually affects how much meaningful learning takes place in a classroom.

Opinions about standardized tests influence on teaching and learning process are conflicting. Smith et al. (1989) found that pressure to improve students' test scores caused some teachers to "neglect material that the external test does not include reading real books, writing in authentic context, solving higher-order problems, creative and divergent thinking projects, longer-term integrative unit projects, [and] computer education..." (p. 268). She and her colleagues also found that teachers were using worksheets with the same question format as the mandated test [⁵<https://www.academia.edu/71353844/EffectsofStandardizedTestingonTeachingandLearningAnotherLook>].

Teachers begin to start "teaching of tests" making the classroom a narrower educational experience and thus limiting the subject matter discussed in class. Since most of these tests are multiple choice limits teaching and learning to knowledge, at the expense of skills and abilities, such as critical thinking, creative thinking, and problem solving. One very large qualm about standardized tests is that they do not take in diversity enough. Gender, socioeconomic classes, culture, language difference might be a barrier for a student to succeed on a test.

Besides, they show the final answer, omitting the mind process of the students that led to this solution, the personal abilities and attitudes of every student cannot become evident from the test alone since the test stresses only on certain characteristics. From the short overview of the advantages and disadvantages of the popular test

method in the recent years, we can conclude that this process is suitable for checking and evaluating of the knowledge and the skills of the students on every level of competency only if combined with other methods of evaluation [Stewart p.33, 1998].

Standardized tests can't measure initiative, creativity, imagination, conceptual thinking, curiosity, effort, irony, judgment, commitment, nuance, good will, ethical reflection, or a host of other valuable dispositions and attributes. What they can measure and count are isolated skills, specific facts and function, content knowledge, the least interesting and least significant aspects of learning [[6https://newlearningonline.com/new-learning/chapter-9/kohn-on-standardised-tests](https://newlearningonline.com/new-learning/chapter-9/kohn-on-standardised-tests)].

To a great extent a non-standardized test is marked with its flexibility in terms of scope and format, variable in difficulty and significance. Non-standard tests, in comparison to standard tests are usually developed by individual instructors, the format and difficulty of these tests may not be widely adopted or used by other instructors or institutions. The usage and purposes of non-standard test may vary as well. A non-standardized test may be used to determine the proficiency level of students, to motivate students to study, and to provide feedback to students. In some instances, a teacher may develop non-standardized tests that resemble standardized tests in scope, format, and difficulty for the purpose of preparing their students for an upcoming standardized test. Finally, the frequency and setting by which a non-standardized tests are administered are highly variable and are usually constrained by the duration of the class period [[7https://www.academia.edu/29689342](https://www.academia.edu/29689342)].

Both standardized and non-standardized tests usually consist of a series of items. An item is a specific task to perform and can test one or more points or objectives. For example, an item can test one point such as the meaning of the given vocabulary word or several points such as an item, which tests the ability to obtain facts from a passage and then make inferences based on these facts. Likewise, a given objective may be tested by a series of items. For example, there could be five items all testing one grammatical point: tag questions [Cohen p.381, 2000].

Items can be more discrete or more integrative in nature just as they can be more objective or subjective. A completely discrete point item would test simply one point or objective at a time. Sometimes an integrative item is really more a procedure than an item, as in the case of free composition which could test a number of objectives.

The objectivity of an item refers to the way it is scored. A multiple choice item is objective in that there is only one right answer [Cohen p.382, 2000]. A free composition may be more subjective in nature if the scorer is not looking for any one right answer, but rather a series of factors including say creativity style cohesion and coherence grammar and mechanics.

Below we provide the advantages and disadvantages of both types of tests.

Analysis

Standardized tests have the following advantages:

- 1) They are perceived to be more fair and more reliable in comparison to non-standardized tests, there is no need for teachers to have their bias or emotions to influence the score of the students.
- 2) Tests are scored objectively, machines do the grading for these tests and computer track the progress of achievement for different groups of students in different subjects.
- 3) These tests provide the opportunity for self-study and self-evaluation.
- 4) One of the most important advantages of testing is the feedback and evaluation of academic progress. The data gathered from assessment results enables governing bodies to determine areas of proficiency and gaps in student learning and instigate change or reform.
- 5) There is also an accountability factor for students, teachers and everyone involved in the educational system.

The shortcoming of standardized tests are as follows:

- 1) Standardized testing affects the curriculum being taught in schools. Teachers begin to start "teaching of tests" making the classroom a narrower educational experience and thus limiting the subject matter discussed in class.
- 2) Non-standardized tests can't measure initiative, creativity, imagination, conceptual thinking, curiosity, effort, judgment, commitment, nuance, good will or a host of other valuable dispositions and attributes.
- 3) They can measure and count separated skills, specific facts and function, content knowledge, the least interesting and least significant aspects of learning.
- 4) And commonly standardized-tests contain bias.

5) Communicative skills are suppressed.

Non-standardized tests have the following advantages:

- 1) A non-standardized test is usually flexible in scope and format, variable in difficulty and significance.
- 2) Standardized tests measure initiative, creativity, imagination, conceptual thinking, curiosity, effort, judgment, commitment, nuance, good will and a host of other valuable dispositions and attributes.
- 3) Non-standardized tests require more time.
- 4) They allow the typical mistakes and difficulties of students to be discovered.
- 5) These tests affect the curriculum being taught in schools.

The lack are as follows:

- 1) These tests are not scored objectively, sometimes teachers have their own bias or emotions to influence the score of the students.
- 2) These tests do not provide the opportunity for self-study and self-evaluation.
- 3) The communicative and speaking skills are suppressed.

Conclusion

Thus, we can conclude that both standardized and non-standardized tests are used to assess student's academic progress, diagnose an analysis of student weaknesses and strengths, measure student's ability to acquire knowledge and place students in appropriate instructional programs. These tests can also be used to provide information for program evaluation and improvement. Teachers play a central role in tests and testing assessments. They are integral in test selection, student preparation, test administration, and interpreting and communicating results to students and their caregivers.

References:

1. Cohen, A. *Second Language Testing*. Hillsdale: Erlbaum publication, 2000.
2. Celce-Murcia, M. *Teaching English as Second or Foreign Language*. Chicago: University, 2001.
3. Genesee, F. and Upshur, J.A: Classroom-based evaluation in second language education. Cambridge: Cambridge University Press. 1996.
4. Goh, D. *Exert from Assessment Accountability for Diverse Learners*. Indiana:
5. Indiana University Press, 2004.
6. Higgs, J., John, T. *Testing Language Skills*. Orlando: Harcourt Brace Jovanovich, 1991.
7. Hamp-Lyons, L. *Applying ethical standards to portfolio assessment of writing English as a second language*. In Milanovic, M. and Saville, N., editors, Performance testing, cognition and assessment. Cambridge: University of Cambridge, 1996
8. Lewkowicz, J.A.: *Authenticity in language testing: some outstanding questions*. *Language Testing*, 2000.
9. Stewart, H. *The Student Evaluation Standards*. Newbury Park, CA: Ablex Publishers corporation, 1998.
10. Thissen, D., Waizer, H. *Test Scoring*. Mahwah: Newbery House, 2001.
11. Wigglesworth, G, *The effect of planning time on second language test discourse*. In Kunnan, A.J., editor, Validation in language assessment. Mahwah, NJ: Lawrence Erlbaum, 1998.
12. <https://files.eric.ed.gov/fulltext/EJ1246414.pdf>
13. <https://methods.sagepub.com/reference/the-sage-encyclopedia-of-educational-research-measurement-and-evaluation/i19548.xml>
14. <https://methods.sagepub.com/reference/the-sage-encyclopedia-of-educational-research-measurement-and-evaluation/i16318.xml>
15. https://www.academia.edu/38027992/Standardized_and_Non_Standardized_Tests_STANDARDIZED_AND_NON_STANDARDIZED_TESTS_Non_Standardized_Tests
16. https://www.academia.edu/71353844/Effects_of_Standardized_Testing_on_Teachers_and_Learning_Another_Look
17. <https://newlearningonline.com/new-learning/chapter-9/kohn-on-standardised-tests>
18. <https://www.academia.edu/29689342>

Получено: 29.09.2022

Received 29.09.2022

Рассмотрено: 06.10.2022

Reviewed 06.10.2022

Принято: 14.10.2022

Accepted 14.10.2022



This work is licensed under a Creative Commons Attribution-Non Comercial 4.0 International License