

THE ROLE AND IMPORTANCE OF TESTS IN THE ASSESSMENT OF KNOWLEDGE.

Uchkurova Zarina Shavkatovna.
Teacher , Department of Zoology and Anatomy Natural sciences,
Tashkent State Pedagogical University named after Nizami
Annotation

This article dedicates to the role and importance of tests in the assessment of knowledge and technology of students' achievement in the area of educational activities, methods, techniques, forms and conditions of monitoring knowledge quality in accordance with the requirements of Uzbek higher education system modernization. The author propose methodic techniques of students' training for testing based on innovative and traditions approaches, development of the system of student's competences evaluation in the course of teaching using current educational technologies in higher institutions. Pedagogical testing is represented as a set of instruments that allows assuring systematic monitoring and objective evaluation of students' professional training quality in the course of entire education process, defining stages of tests carrying out and also ways and means of educational results adjustment. The expected result is creation of the organization model of the system of students' knowledge evaluation based on elaborated vocationally-orientated test materials that evaluate competences included in the requirements of State Educational Standards of Higher Education of the third generation.

Key words: test, technology, testing, monitoring, knowledge, evaluation, assessment.

I think that probably the most important thing about our education was that it taught us to question even those things we thought we knew.

To say you have to be testing your knowledge all the time in order to be more effective in what you are doing.

Thabo Mbeki.

Introduction

It is clear that, teaching is a process, which facilitates learning. On the other hand, learning is a mental process, which results in the acquisition of knowledge. It is also vital to note that teachers use tests to evaluate students' performance and see where they can adjust for the sake of the same learner who wants to run away from being examined by the teacher.

Testing is important in all education system because it helps teachers and the pupils or students to determine how much they have taught and learnt, respectively.

It is because of the teacher trying to find the areas of difficulty in order to take corrective measures that tests are administered in university. A teacher who fails to administer any test to the student is doing a disservice to the child. Similarly, a student who keeps away from tests is also doing himself or herself a disservice.

Review literature

Pedagogical science has a whole bunch of tools and methods, which can contribute to achievement of modern objectives of education. Many researchers' works are concerned with theoretical development of the problem of students' professional training quality control assurance in higher institutions. Testing and assessment are of major importance for education because assessments and tests determine to a large extent what students learn, how they plan their studies, and the effort they expend (Bishop, 2002; Crooks, 1988; Miller & Parlett, 1974; Popham, 2003). The form and content of tests further influence how students study and how they perceive the nature of the courses and the profession (Entwistle, 1996; Gielen, Dochy, & Dierick, 2003; Snyder, 1971).

In higher education, many courses are concluded with achievement tests to measure the students' attainment. Therefore, many teachers in higher education have to develop such end-of-course exams (Anderson, 1987; Mavis, Cole, & Hoppe, 2001; Parkes, Fix, & Harris, 2003). Often, tests developed by teachers are referred to as in-house or teacher-made tests (Jozefowicz et al., 2002).

Today, a growing number of teachers provide students with digital online activating learning materials or quizzes that lead up to these end-of-course exams (Nicol, 2007). For both purposes, especially for large enrolment courses in undergraduate education, quizzes and tests consisting of selected response test item formats (such as true-false, multiple choice, matching, ordering, drag-and-drop, and numeric) are developed by teachers to keep grading within reasonable time limits and be able to cover a fairly broad number of topics in a short time span. In this article, these selected response test items, whether hard copy or electronic, will be called simply *test items*.

Individual test items form the core building blocks of these tests and the test items determine the quality of those tests; in the words of Haladyna (1997): "Good tests consist of good test items." It is the test item that determines which cognitive processes are elicited, whether more or less informed students are appropriately identified, and students' perceptions of the nature of the subject domain.

Further, designing test items is a *recurring* task for teachers in higher education for two main reasons. First, especially in higher education, the content of courses changes often due to new insights in the domain or curricular considerations, so new test items need to be devised. Second, because tests are often made public after having been administered to provide students with opportunities to learn or to allow them to check the correctness of the scoring, those questions become unfit for use in subsequent tests, forcing teachers to produce test items again and again.

It has been reported that test items developed by teachers in higher education suffer from low quality (DiBattista & Kurzawa, 2011; Hansen & Dexter, 1997; Holsgrove & Elzubeir, 1998; Jozefowicz et al., 2002). Sometimes teachers are not aware of common construction errors or are not able to avoid common construction errors. Such errors can confuse student leading them

astray in answering correctly or – in contrast – such errors can unintentionally lead students to the correct answer without them knowing the materials. But also, teachers may solely design test items that merely quote information verbatim from instructional materials or only call for recognition or recall of information without any context and therefore not for ‘higher-order thinking’ or low levels of cognitive demand. Querying for higher-order thinking or learning outcomes such as application of knowledge, insight or problem solving is a need often expressed by teachers and learners. It is reported however that though teachers do try to query for high erorder learning outcomes such as critical thinking or problem solving, they are not able to design such items successfully and thus fall back on test items that indeed merely elicit recall of factual information (Buckles & Siegfried, 2006). Test items that only elicit recognition of factual information encourage students to use surface learning approaches (Trigwell & Prosser, 1991; Vermunt, 2005) in which pure memorization and rote learning are most effective (Beattie, Collins, & McInnes, 1997; Boud & Falchikov, 2007; Smith & Miller, 2005; Struyven, Dochy, & Janssens, 2003; Vermunt, 2005). Though for certain subjects such as language learning or anatomy, this kind of knowledge acquisition and testing may be appropriate, such studying strategies are considered inappropriate for higher education in general (Biggs, 1996; Boud & Falchikov, 2007; Draper, 2009; Entwistle & Entwistle, 2003; Struyven et al., 2003). Tests that contain solely test items that query decontextualized information most probably reinforce the low appreciations of students and the general public regarding the value of selected response test items (for example Beattie et al., 1997; Cizek, 2001) and the conviction that selected response test items cannot query for high cognitive demand.

For that matter however, Rodriguez (2003) argued that it is in particular the *stimulus* of the test item that is of most importance when thinking about cognitive demand. The review study of Rodriguez suggested that correlations between the scores on selected response test items (multiple choice answering mode) and constructed response test items (open ended question answering mode) were high when the stimulus (the task or problem presented) of the two different response modes were comparable. Therefore, generating problems or tasks with stimuli aiming at querying for understanding, critical thinking or problem solving requiring higher cognitive demand is essential for selected response test items to fullfill that role.

In this article, such higher-order learning outcomes are operationalized as outcomes in which learners need to demonstrate that they have knowledge and skill to *use* learnt facts, concepts, principles and procedures in *familiar but new* situations within the bounds of a course content. In this article, such new situations are considered to be test questions in which information has been reframed, added or excluded, or in which new problems, contexts or answering options are offered. Such new information requires the student to reevaluate, reinterpret and transform the learnt materials in order to deal with these new

situations that are presented in the stimuli and answering options of the test item. The level of cognitive demand in those circumstances is a function of the examinees stance on the continuum from novice to expert. This continuum (novice-expert) is broad so the cognitive demand is never really known unless one questioned the examinee. Higher-order thinking in this article is thus related to the learner's ability to show *transfer* (Bransford & Schwartz, 1999; Chi, Glaser, & Farr, 1988; Wiggins & McTighe, 1998) of knowledge and skill to comparable, yet different propositions, problems or situations. It is therefore vital for higher education that teachers be able to generate a range of test items that are appropriate for the variety of goals for which test items are needed. Teachers should be able to design test items not only for rote learning purposes but also and more importantly for querying higher-order learning goals such as application of knowledge, critical thinking, and problem solving.

Writing good-quality test items, however, is not a straightforward or easy task to execute. It has been reported that developing high-quality test items takes

substantial amounts of time (Case, Holtzman, & Ripkey, 2001) and training (Downing & Haladyna, 2006); the task is regarded as difficult, especially in terms of developing sufficiently numerous plausible and distractors (Mayenga, 2009).

In practice of course, teachers are not specifically equipped and trained to design test items. Typically, teachers in higher education work alone when generating test items and rarely seek or receive support (Cizek, Fitzgerald, & Rachor, 1995; Conole & Warburton, 2005; DiBattista & Kurzawa, 2011; Downing, 2002; Jozefowicz et al., 2002; Popham, 2003; Sahari, 1999; Van der Vleuten et al., 2004). In practice, teachers are not quick to use an item-writing guide or handbook. Further, in formal one-time learning events such as workshops or training, the information obtained or skills learned can be forgotten if there is no context in which they are shared, fostered, and put into practice (Knight, Tait, & Yorke, 2006).

Optimal solutions to raise the expertise of teachers, the quality of test items, and the process of test item design entail more flexible approaches and organizational contexts such as communities of practice, peer review, and on-the-job training (Knight et al., 2006; Wenger, 1998). However, in the realities of higher education, instituting such learning contexts for teachers is seldom practical or affordable (Iramaneerat, 2012; Warburton, 2009). Thus, the present situation leads to test items in higher education produced in large quantities with the weaknesses reported above, which is highly problematic for ensuring the quality of assessment and learning.

Results

Student's knowledge quality testing is innovative technology of education quality evaluation, which expresses quality of conformance to federal state educational standards. Testing is one of knowledge control forms in the framework of the institutional system of students' training quality and is used along with other types and forms of control. It is testing that can

motivate students' active cognitive and creative activity for the whole period of education process. Various kinds of pedagogical testing of knowledge and skills are used in the system of higher education: entrance testing (at the beginning of education); current testing (during education), borderline testing (upon topic completion), final testing (upon subject completion), delayed testing (control of "residual" knowledge some time after topic or course studying). Tests used for working with TSPU named after Nizami students for the purpose of knowledge control are diverse and include: cross choice (selection of pairs from two blocks by some or other features); alternative choice (selection of one answer of the offered); multiple choice (selection of one answer of the three-four offered); ordering (making a connected text of fragmented paragraphs, sentences, placing them in a logical sequence or a sentence of words); conclusion/ending (free answer or formulation of student's own evaluation); replacement/substitution of particular words, phrases, dates, etc.; transformation of proposed expressions, sentences, etc.; answers to questions; close-test (reconstruction of missing words in a text), such test can be based on multiple choice or it may not have it at all, etc.

Gained experience of pedagogic testing revealed both benefits and drawbacks in the process of use for test measuring materials knowledge quality evaluation. The benefits are the following: substantial time savings (it takes one hour to assess a group of 30-32 people while common exams take 6-10 hours per group); savings on high-skilled teachers' salaries (testing can be carried out without teachers); high objectivity and consequently weakening subjective factor of teacher's personal attitude; uselessness of "cribs"; higher differentiation of grades that allows detecting completeness of students' knowledge, and not its part when answering tickets; emergence of students' self-education incentive, strengthening their own role in education; easing of emotional tension (the examinee feels less stress than if he or she answers orally); test system allows the higher education institution administration to monitor education process advancing and increases teachers' responsibility for their performance. Certainly, testing as a control method has drawbacks: the student cannot state his thoughts extensively, with complete sentences; testing system of knowledge assessment doesn't make the student think in a complex manner, to separate the main from the secondary; tests results depend on many side factors (health, motivation for result, etc.); time limitation for accomplishing test causes nervousness and prevent some students from keeping their mind on the correct answer, even if they know it; there's possibility of technical error, ill-defined formulation of separate test tasks, potential danger of test base hacking, etc.

High technologies allow monitoring education process in the form of computer testing carried out for the purpose of independent unbiased information acquisition regarding the level of learners' mastery of subject, conformance of content, level, and quality of training to State Educational Standards of Higher Education (FSES HPE) of the third generation for bachelors and masters under institutional system of education quality control; shaping students' skills regarding working with tests; the use of testing results in the course of interim and final attestation. Within the framework of the project "Internet testing" and "Internet trainers" the technology of Internet exam in the area of professional education is applied in Tashkent State Pedagogical University named after Nizami. These technologies are a software system, in which original techniques of knowledge assessment and goal-oriented training of learners in the course of multiple passing of tests of varying complexity are taken as a basis. Participation in this project allows the higher institution to independently prepare students for external procedures of knowledge quality control (competence-based and traditional approaches in the field of professional education); accreditation testing; unscheduled control and supervision procedures), and also carry out internal control of knowledge quality. On the basis of SES HE requirements the model of pedagogical measure is represented in three interrelated units. The first unit of tasks examines student's mastery of subject material at the level "to know". This unit includes tasks where method of solution learned by the student during subject studying is obvious. Tasks of the second block assess level of mastery of subject material at the level "to know" and "to be able". This unit includes tasks where there is no explicit guideline for solution and the student himself chooses one of learned methods for its solution. The third unit evaluates mastery of subject at the level "to know", "to be able", and "to master". It's represented by cases, the content of which suggests the use of a complex of skills so that the student could develop the way of solution, combining methods he knows and attracting knowledge from other subjects. In his turn, the student has access to extensive information regarding all test sessions passed as well as possibility to track all testing stages planned for him by teacher. The student logs in using personal number of his grade book.

The results which students achieve during the experiment it shows this statement:

The content of experimental and control group work

<u>No</u>	<u>Educational establishment</u>	<u>Experimental group</u>	<u>Amount of students</u>	<u>Control group</u>	<u>Amount of students</u>	<u>Total number of the students</u>
	<u>Tashkent State Pedagogical University</u>	302	30	301	32	62

The results of the experimental group checked according to the State Education Standard

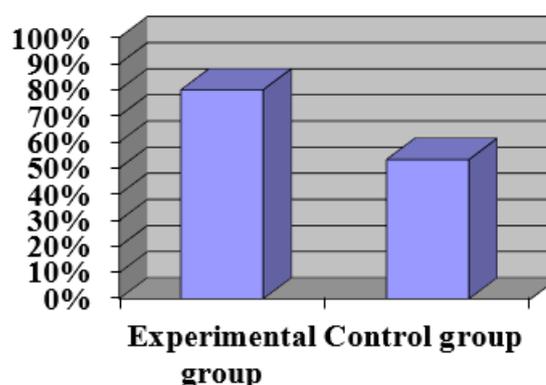
<u>№</u>	<u>Groups</u>	<u>Amount of students</u>	86-100%	71-85%	55-70%	0-54%	<u>Average</u>

The results of the control group checked according to the State Education Standard

<u>№</u>	<u>Groups</u>	<u>Amount of students</u>	86-100%	71-85%	55-70%	0-54%	<u>Average</u>
1	301	32	5	17	10	-	68,75%

	<u>Experimental group</u>		<u>Control group</u>
<u>Excellent</u>	9	39,00%	
<u>Good</u>	17	47,00%	
<u>Satisfactory</u>	5	35,00%	

<u>Average</u>	<u>Experimental group</u>	<u>Control group</u>
100	86,67%	68,75%



Re

1. Ac
2. D. Mumford (Ed.), *Handbook of organiza*
San Diego, TX: Academic Press.
3. Anderson, S. B. (1987). The role of the teacher-made test in nigner educauon. *New Directions for Community Colleges, 1987(59)*, 39–44.
4. Barab, S., & Squire, K. (2004). Design-based research: Putting a stake in the ground. *Journal of the Learning Sciences, 13(1)*, 1–14.
5. Beattie, V., Collins, B., & McInnes, B. (1997). Deep and surface learning: a simple or simplistic dichotomy? *Accounting Education, 6(1)*, 1–12.
6. Becker, S. A., & Berkemeyer, A. (2002). Rapid application design and testing of web usability. *IEEE Multimedia, 9(4)*, 38–46.
7. Biggs, J. B. (1996). Enhancing teaching through constructive alignment. *Higher Education, 32(3)*, 347–364. <http://doi.org/10.1007/BF00138871>
8. Bishop, J. H. (2002). What is the appropriate role for student achievement standards? In *Conference Series-Federal Reserve Bank of Boston* (Vol. 47, pp. 249–278). Federal Reserve Bank of Boston; 1998. Retrieved from
9. Boud, D., & Falchikov, N. (2007). *Rethinking assessment in higher education: Learning for the longer term*. Routledge.
10. Bransford, J. D., & Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with multiple implications. *Review of Research in Education, 24*, 61
11. Buckles, S., & Siegfried, J. J. (2006). Using multiple-choice questions to evaluate indepth learning of economics. *The Journal of Economic Education*.
12. Buxton, B. (2010). *Sketching User Experiences: Getting the Design Right and the Right Design*. Morgan Kaufmann.
13. Case, S. M., Holtzman, K., & Ripkey, D. R. (2001). Developing an item pool for CBT: A practical comparison of three models of item writing. *Academic*

- Medicine*, 76(10), S111–S113.
14. Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, 5(2), 121–152. http://doi.org/10.1207/s15516709cog0502_2
 15. Chi, M. T. H., Glaser, R., & Farr, M. J. (Eds.). (1988). *The nature of expertise*. Hillsdale, NJ: Lawrence Erlbaum Associates.
 16. Cizek, G. J. (2001). More unintended consequences of high-stakes testing. *Educational Measurement: Issues and Practice*, 20(4), 19–27.
 17. Cizek, G. J., Fitzgerald, S. M., & Rachor, R. A. (1995). Teachers' assessment practices: Preparation, isolation, and the kitchen sink. *Educational Assessment*, 3(2), 159–179.
 18. Conole, G., & Warburton, W. (2005). A review of computer-assisted assessment. *ALT-J*, 13(1), 17–31.
 19. Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, 58(4), 438–481.
 20. Cross, N. (2007). Forty years of design research. *Design Studies*, 28(1), 1–4.
 21. Crouch, C. H., & Mazur, E. (2001). Peer instruction: Ten years of experience and results. *American Journal of Physics*, 69(9), 970–977.
 22. Davidson, J. E., & Sternberg, R. J. (2003). *The psychology of problem solving*. Cambridge University Press.
 23. Dembo, M. H., & Howard, K. (2007). Advice about the use of learning styles: a major myth in education. *Journal of College Reading and Learning*, 37(2), 101
 24. DiBattista, D., & Kurzawa, L. (2011). Examination of the quality of multiple-choice items on classroom tests. *The Canadian Journal for the Scholarship of Teaching and Learning*, 2(2), 4.
 25. Downing, S. M. (2002). Threats to the validity of locally developed multiple-choice tests in medical education: Construct-irrelevant variance and construct underrepresentation. *Advances in Health Sciences Education: Theory and Practice*, 7(3), 235–241.
 26. Downing, S. M., & Haladyna, T. M. (Eds.). (2006). *Handbook of test development*. Mahwah, NJ: Lawrence Erlbaum Associates.
 27. Draper, S. W. (2009). Catalytic assessment: understanding how MCQs and EVS can foster deep learning. *British Journal of Educational Technology*, 40(2),
 28. Entwistle, N. (1996). Recent research on student learning. In J. Tait & P. Knight (Eds.), *The Management of Independent Learning* (pp. 67–112). London
 29. Entwistle, N., & Entwistle, D. (2003). Preparing for examinations: The interplay of memorising and understanding, and the development of knowledge objects. *Higher Education Research and Development*, 22(1), 19–41.
 30. Gielen, S., Dochy, F., & Dierick, S. (2003). Evaluating the consequential validity of new modes of assessment: The influence of assessment on learning, including pre-, post-, and true assessment effects. *Optimising New Modes of Assessment: In Search of Quality and Standards*, 37–54.
 31. Gierl, M. J., & Haladyna, T. M. (2012). *Automatic item generation: theory and practice*. New York, NY: Routledge.

32. Gronlund, N. E. (1998). *Assessment of student achievement*. Boston, MA: Allyn & Bacon.
33. Haladyna, T. M. (1997). *Writing test items to evaluate higher order thinking*. Needham Heights: Allyn & Bacon.
34. Haladyna, T. M. (2004). *Developing and validating multiple-choice test items* (3rd Edition). London: Lawrence Erlbaum Associates.
35. Hansen, J. D., & Dexter, L. (1997). Quality multiple-choice test questions: Item writing guidelines and an analysis of auditing testbanks. *Journal of Education for Business*, 73(2), 94–97.
36. Hattie, J., Jaeger, R. M., & Bond, L. (1998). Persistent methodological questions in educational testing. *Review of Research in Education*, 24, 393–446.