

Turkish Prospective Teachers' Perspective of Different Types of Exams: Multiple Choice, Essay and Computerized-type Testing

Neşet Demirci
Balıkesir University

Abstract

The major aim of this study was to compare prospective teachers' attitudes and natures toward teacher-made essay, multiple-choice type versus computerized-type exams. The primary study was conducted on a sample of 393 prospective teachers (specifically, students from 33 physics education, 93 science education, 66 computer education, and 201 elementary education departments) at Necatibey Faculty of Education, in Balıkesir, Turkey who were administered a test attitude inventory specifically designed to assess prospective teachers' attitudes toward essay, multiple choice versus computerized type formats on a variety of critical dimensions. The data from study was pointing to more favorable prospective teachers' attitudes towards multiple choice exams compared to essay and computerized- type formats on most dimensions assessed. However, prospective teachers, in general, did not want to choose one type to another; because, they are willing to use some assessment types altogether or combination of at least two types (multiple choice and essay). Many see the computerized-type exam is more contemporary approaches than the others, and also many have a positive attitude toward using it in their further teaching. Nevertheless, somehow many see using computerized-type exam is not convenient and/or comfortable to use it yet.

Introduction

Assessment is one of the essential aspects of all instruction to maximize opportunities for meaningful learning (Carless, Joughin & Mok, 2006). Evaluation refers to a more formal mode of assessment and is the basis for judging the skills or knowledge of the students as well as the effectiveness of a unit or activity (New Horizons for Learning, 2002). Assessment can take the form of a quiz or examination to test students' learning achievements, or of a questionnaire to investigate students' attitudes and reactions to instructional courseware. Instructors have to know what and how well students have learned, and so do students themselves. As classroom testing experts (for examples, Gronlund, 1976; Thorndike and Hagen, 1969) have pointed out, the choice of a particular item format should normally be determined by theoretical as well as practical considerations, such as: the relative ease with which various test objectives are measured; the degree of difficulty in constructing or scoring items; freedom from irrelevant sources of variation in test results; degree of precision required in reporting results; and so on.

The item formats most often used in the construction of classroom tests may be conveniently classified into two broad categories (Gronlund, 1976): the more objective and structured selection type formats (e.g. multiple choices, true/false, matching, etc.), requiring the examinee to select the correct answer among a number of given alternatives, and the more subjective construction type format (e.g., essay, short answer), permitting the examinee to organize, construct, and present the answer in written form. A distinct difference in learning approaches according to assessment type was observed by Scoullar (1998), Ramsden (1988), and Watkins (1982). These researchers noted that students perceived the essay assignment as assessing high levels of cognitive processing and were more inclined to employ both deep strategies and motives when preparing for their essay than when preparing for their multiple-choice examination. Students perceived the multiple-choice examination as assessing ability to recall factual information (lower levels of cognitive processing) and were more inclined to

employ surface strategies and motives when preparing for such an examination than when preparing for an essay assignment.

Two types of examination that students commonly encounter in their study are objective (for example, multiple-choice) and essay examinations (Smith and Miller, 2005). Biggs (1993) and Claxton and Murrell (1987) each maintained that in an objective tests students are examined on relatively specific information that is predominantly knowledge-based, where the learner is asked to give a specific answer or to select the correct response from alternatives provided. Biggs argued that in a multiple-choice examination, students are encouraged to adopt a convergent type of strategy whereby factual information and details are focused upon. This form of test, in Biggs' and Entwistle (1996)'s view, encourages students to rote-learn (with minimal emphasis on understanding) in order to maximize accurate recall of the information learned. Multiple-choice, therefore, can be seen to influence the learner to adopt a surface approach in learning.

The traditional perception is that Multiple Choice Questions (MCQs) can only be used for testing lower level cognitive skills. This is not true; according to Hibberd (1996) "... they can be implemented to measure deeper understanding if questions are imaginatively constructed." Students were more likely to employ deep learning approaches when preparing their assignment essays which they perceived as assessing higher levels of cognitive processing. Poorer performance in the assignment essays was associated with the employment of surface strategies (Scouller, 2000).

There are many advantages to multiple-choice testing. Although difficult to create, they are easy to score and therefore are the evaluation method of choice in large classes. The added benefit is that taking a test generally improves students' performance on a later test; this is referred to as the testing effect (Bjork, 1975; Carrier & Pashler, 1992; Glover, 1989; Hogan & Kintsch, 1971; Izawa, 1970; Kuo & Hirshman, 1996; McDaniel & Masson, 1985; Runquist, 1986; Spitzer, 1939; Thompson, Wenger, & Bartling, 1978; Tulving, 1967).

The different thoughts outlined and discussed in the literature for choosing one item format over another in planning a classroom test generally revolve around three major factors of concern in the test endeavor: a) the subject matter domain assessed (for examples; adequacy and ease of measuring specific course objectives); b) the test constructor or user (for examples; ease of test preparation, ease of scoring tests, etc.); and various extraneous factors (for examples; guessing, copying, bluffing) possibly affecting the psychometric properties of the test scores. However, when planning a classroom test, one major factor is the perspective of the student examinee taking the test. Which particular format do students perceive to be more convenient, interesting, motivating, anxiety evoking, eliciting greater success expectancies and so on? These and other questions have not been sufficiently addressed in school-based evaluation research, with classroom testing experts generally paying little attention to the examinees' perspective-one of the most potentially useful sources of information about the subjective qualities of the test or its constituent components.

Given the assumption that examinees, are one of the best sources of information about the subjective qualities of a test (or its constituent components), and that examinees' test attitudes and dispositions should be taken into consideration by test constructors and users when deciding upon test construction and administration policy (Nevo, 1985, 1986; Zeidner, 1985, 1986, 1987), it is truly surprising that so little research has been devoted towards assessing examinees' attitudes toward varying facets of classroom testing. Furthermore, very

little work has been devoted towards the development and implementation of specific feedback systems designed to study examinees' reaction towards various facets of the classroom test.

Computerized-testing

Computer use is shaping the ways tests are constructed, normed, validated, administered, and scored. The reach of this technology on test practices is broad and international. Thus, involvement of organizations that transcend one community, even one nation, is needed to envision, create, promote, regulate, and in other ways assist in forming and reforming test-related services that serve the professions responsible for test development and use as well as the public.

With the advent of the personal computer, the development of computer-administered versions of paper and pencil tests is shown rapidly. These provided some advantages over paper and pencil, in terms of control of administration, and some disadvantages (for example, the need for sufficient hardware to test groups of people). They also raised the question of equivalence with their paper and pencil counterparts. Most research (for examples, Bartram, 2005; Mead & Drasgow, 1993) has tended to show that equivalence was not a key problem so long as the tests were not speeded measures of ability.

Despite the increasing sophistication of computer-based assessment systems, within the field of occupational assessment the tests they contain are, typically, computer implementations of old paper-and-pencil tests. Nevertheless, there has been innovation in the field and the consequences of that innovation are increasingly finding their way into commercial practice. Tests can be pioneering in a number of different ways. The most obvious is where the actual test content is innovative. However, innovation can also occur in less obvious ways. The process used to construct the test may be innovative and rely on computer technology and the nature of the scoring of the items may be innovative. In practice there is an interaction between these different aspects of innovation, in that some of the most interesting developments in test content also involve innovation in how that content is created (Bartram & Hambleton, 2006).

The use of computer-based testing is increasing rapidly. It has been helped not only by the development of better interfaces, but by the spectacular increases in volume of and accessibility to hardware. More than 50 new item types have been reported (Hambleton & Pitoniak, 2002; Zenisky & Sireci, 2002), with many more variations on the way. Drasgow and Mattern (in Bartram and Hambleton, 2006) offer many item type variations: they may involve complex item stems, sorting tasks, interactive graphics, the use of both audio and visual stimuli, job aids (such as access to dictionaries), joy sticks, touch screens, sequential problem-solving, pattern scoring, and more.

Already, we are seeing hundreds of credentialing agencies in the US delivering their tests at a computer, and many more are coming on board. Admissions tests such as the Graduate Record Exam (GRE), the Test of English as a Foreign Language, better known as TOEFL, and the Graduate Management Admissions Test (GMAT) are now on-line. Maybe, the biggest change in the next decade will likely be the administration of more tests at a computer (for example, Luecht, 1998; Luecht & Clauser, 2002; Wainer et al., 2000; van der Linden & Glas, 2000).

Obviously, many new item types can be expected in the future since there is little point to assessing examinees at a computer and not taking advantage of all of the valid options that are available with computer-based testing. A consideration of computer-based tests in accounting, architecture, and medicine highlight what is possible if substantial funds and time are available (see, for example, van der Linden & Glas, 2000; Irvine & Kyllonen, 2002; Pitoniak, 2002). However, even without substantial funds, improvements in assessment practices are possible with new item types involving sorting, classifying, and ranking tasks, and with automated scoring software, more frequent use of sentence completion, short answer, and extended answers. A review of the new item types described by Drasgow and Mattern (in their chapter—see Bartran and Hambleton, 2006), and by Zenisky and Sireci (2002), clearly points to new item types that can be implemented relatively easily, with little cost, and with increasing test validity.

Merrell and Tymms (2007) described the development of an adaptive assessment called Interactive Computerized Assessment System (InCAS) that is aimed at children of a wide age and ability range to identify specific reading problems. Rasch measurement has been used to create the equal interval scales that form each part of the assessment. The rationale for the structure and content of InCAS is discussed and then different formats of feedback supplied to teachers are explained. This feedback is accompanied by research-based strategies for remediation, following the principle of assessment for learning on how to improve.

Olsen et al. (1986) compared paper-administered, computer-administered, and computer-adaptive tests by giving third- and sixth-grade students mathematics applications achievement tests. This study found no significant differences between paper-administered and computer-administered tests, and equivalences among the three test administrations in terms of score rank order, means, dispersions, and distribution shapes.

In view of the gaps in the classroom testing and evaluation literature, the major aim of the present study is twofold: a) to systematically compare and contrast the preferences, attitudes, and perceptions of students examinees with respect to test formats currently in use for constructing teacher-made tests, namely, essay, multiple-choice types and computerized type exams; and b) to describe the characteristics, potential use, and application of a test attitude inventory (adapted from Zeidner (1987)'s study), specifically designed to gather data on examinees' attitudes towards varying item formats.

Methodology

Subjects

The sample consisted of 393 volunteered undergraduate prospective teachers from Necatibey Educational Faculty, Balikesir University, situated in North-west of Turkey. The entire sample was distributed almost equally by sex (male, 48.1 % female, 51.9 %). The sample consisted of four different departments; specifically, 33 students from physics education, 93 students from science education, 66 students from computer education, and 201 students from elementary education departments. These samples (prospective teachers) volunteered to participation of the study in the educational years of 2006-2007.

Instrument and Procedures

A test attitude inventory (was adopted from Zeidner's (1987) study, added the computerized section—see Table 1) was used for the purpose of gathering data on prospective teachers' perceptions and attitudes towards varying test formats (i.e. multiple-choice, essay

vs. computerized type ones). The inventory consisted of two main parts, briefly descriptions of instruments were given below:

First part:

Likert-type rating scale consisted of 10 Likert-type items, on a five point continuum. Examinees were asked to rate each stimuli, “Multiple-choice type Classroom Test”, “Essay Type Classroom Test” and “Computerized Classroom Test,” separately along the following ten different dimensions. a) *Perceived facility* (5= very easy... 1=very difficult); b) *perceived complexity* (5=not complex at all... 1=very complex); c) *perceived clarity* (5=very clear... 1=very unclear); d) *perceived interest* (5=very interesting... 1=not at all interesting); e) *judged trickiness* (5=not tricky at all... 1=too tricky); f) *perceived fairness* (5= very fair... 1=not at all fair); g) *perceived value* (5=very valuable ... 1=not at all valuable); h) *success expectancy* (5=very high... 1= very low); i) *degree of anxiety evoked* (5=minimal degree of anxiety evoked... 1=high degree of anxiety evoked); and j) *feeling at ease with format* (5=feeling very much at ease... 1=feeling very ill at ease). The stimuli appeared on the inventory in counterbalanced order.

The alpha reliability estimated from Zeliks (1987)’s study, calculated separately for scale ratings of essay, multiple-choice exam, were about .85 in each case, which was considered to be satisfactory for group comparison purposes.

Before using the instrument in the study, the instrument was translated into Turkish language and then it is asked some experts related to understandability and usability of the instruments’ items. After that the first version of the instrument was applied on 92 students from computer education departments. The alpha reliability was calculated as 0,80; 0,83; and 0,81 respectively for each part (essay, multiple choice and computerized essay and multiple choice exam types). According to students’ responses and opinion about instrument’s some items, the slightly changes made on some items of the instrument and then the final version of the instrument was established.

Second part:

The second part of the inventory consisted of a series of relative rating scales, asking prospective teachers to directly compare essay, multiple choice and computer-based exams along the following relevant dimensions, indicating their preference in each case: a) relative ease of preparing for exams; b) reflection of prospective teachers’ actual knowledge; c) technical ease or convenience of usage; d) perceived expectancy of success; e) perceived degree of fairness; f) degree of anxiety evoked by particular test format, and g) overall preference for format. Also, prospective teachers were asked to explain their choice in each case.

The inventory was administered with no set time limit and responded to anonymously by prospective teachers.

Results

This part consisted of two sections; results devoted to first part of the instrument (Likert type scales), and results devoted to second part of the instrument (relative rating scales).

Results devoted to Likert type scales

Table 1 shows prospective teachers’ preferences about different types of exams (the result is given item by item).

Table 1
Prospective teachers' preference about different types of exams

Survey Item-1	Very difficult		Difficult		Normal		Easy		Very easy		No-opinion	
	%	N	%	N	%	N	%	N	%	N	%	N
I rate difficulty of essay type exam as	11.5	45	60.3	237	23.7	93	1.5	6	0	0	3.1	12
I rate difficulty of multiple-choice type exam as	1.5	6	19.8	78	74	290	2.3	9	1.5	6	0.8	3
I rate difficulty of computerized type exams as	16.8	44	32.8	86	29.8	78	9.9	26	3.8	10	6.9	18
Survey Item-2	Very complex		complex		Normal		Little complex		Not at all complex		No-opinion	
I rate complexity of essay type exam as	13.0	51	30.5	120	29.0	134	18.3	72	7.6	30	1.5	6
I rate complexity of multiple-choice type exam as	2.3	9	28.2	111	53.4	210	11.5	45	3.1	12	1.5	6
I rate complexity of computerized type exams as	12.2	32	36.6	96	20.5	80	9.9	26	4.6	12	6.1	16
Survey Item-3	Not at all clear		Little clear		Normal		Clear		Very clear		No-opinion	
I rate clarity of essay type exam as	9.2	36	28.2	111	35.1	138	19.8	78	5.3	21	2.3	9
I rate clarity of multiple-choice type exam as	3.1	12	17.6	69	46.6	183	26.0	102	3.1	12	3.8	15
I rate clarity of computerized type exams as	6.1	16	26.0	68	33.6	88	23.7	62	4.6	12	6.1	16
Survey Item-4	Not at all interesting		Little interesting		Normal		Interesting		Very interesting		No-opinion	
I perceived interest of essay type exam as	27.5	108	30.5	120	24.4	96	11.5	45	3.1	12	3.1	12
I perceived interest of multiple-choice type exam as	11.5	45	21.4	54	46.6	183	14.5	57	2.3	9	3.8	15
I perceived interest of computerized type exams as	8.4	22	13.7	36	30.5	80	29.8	78	10.7	28	6.9	18
Survey Item-5	Too tricky		tricky		Normal		Little tricky		Not at all tricky		No-opinion	
I judged trickiness of essay type exam as	3.1	12	9.9	39	22.1	87	34.4	135	26.7	105	3.8	15
I judged trickiness of multiple-choice type exam as	10.7	42	48.1	189	29.0	114	6.9	27	3.8	15	1.5	6
I judged trickiness of computerized type exams as	6.1	16	17.6	46	26.0	68	23.7	62	16.0	42	10.7	28
Survey Item-6	Not at all fair		Little fair		Normal		Fair		Very fair		No-opinion	
I perceived fairness of essay type exam as	25.2	99	30.5	120	27.5	108	9.9	39	3.8	15	3.1	12
I perceived fairness of multiple-choice type exam as	4.6	18	16.8	66	20.6	81	36.6	144	17.6	69	3.8	15
I perceived fairness of computerized type exams as	11.5	30	14.5	38	29.0	76	32.1	84	5.3	14	7.6	20
Survey Item-7	Not at all valuable		Little valuable		Normal		valuable		Very valuable		No-opinion	
I perceived value of essay type exam as	6.9	27	16.0	63	45.8	180	26.7	105	2.3	9	2.3	9
I perceived value of multiple-choice type exam as	2.3	9	4.6	18	35.1	138	42.0	165	13.0	51	3.1	12
I perceived value of computerized type exams as	3.1	4	8.4	22	41.2	108	32.8	86	9.2	24	5.3	14
Survey Item-8	Very low		Low		Normal		High		Very high		No-opinion	

	%	N	%	N	%	N	%	N	%	N	%	N
I rate success expectancy of essay type exam as	9.9	39	19.1	75	48.9	192	19.1	75	1.5	6	1.5	6
I rate success expectancy of multiple-choice type exam as	2.3	9	9.2	36	48.9	192	35.1	138	3.1	12	1.5	6
I rate success expectancy of computerized type exams as	6.1	16	22.1	58	32.8	86	27.5	72	5.3	14	6.1	16
Survey Item-9	Very high		high		Normal		Low		Very low		No-opinion	
	%	N	%	N	%	N	%	N	%	N	%	N
I rate degree of anxiety evoked of essay type exam as	20.6	81	46.6	183	21.4	84	9.9	39	1.5	6	0	0
I rate degree of anxiety evoked of multiple-choice type exam as	6.9	27	18.3	72	35.9	141	32.1	126	6.1	24	0.8	3
I rate degree of anxiety evoked of computerized type exams as	18.3	48	28.2	74	23.7	62	19.1	50	6.1	16	4.6	12
Survey Item-10	Very ill		Ill		Normal		Ease		Much ease		No-opinion	
	%	N	%	N	%	N	%	N	%	N	%	N
I feel at ease with format of essay type exam as	4.6	18	22.9	90	48.1	189	19.1	75	4.6	18	0.8	3
I feel at ease with format of multiple-choice type exam as	2.3	9	18.3	72	38.9	153	28.2	111	9.9	39	2.3	9
I feel at ease with format of computerized type exams as	4.6	12	31.3	82	35.9	94	15.3	40	6.1	16	6.9	18

Note: About 32% of the all participants did not answer the computerized-type of exam preference items because they did not experienced with computerized type exam.

Examination of category response distribution for the above scales shows that above 49 percent of the sample perceived computerized type of exams to be complex or very complex, in comparison to only about 30 percent similarly perceiving the multiple-choice type and only about 43 percent choose the essay exam. In addition, whereas about 20 percent of the sample judged the multiple-choice exam as being not at all clear or little clear, only about 37 percent felt similarly about essay items, and only about 32 percent felt similarly about computerized type items. Furthermore, multiple-choice type exams were viewed as difficult or very difficult by about 21 percent, whereas 50 percent felt similarly for computerized type, and 71 percent felt similarly for essay exams.

Inspection of category response distribution for the foregoing scales shows that above 40 percent of the sample perceived computerized type of exams to be interesting or very interesting, in comparison to only about 17 percent similarly perceiving the multiple-choice type and only about 15 percent choose the essay exam. In addition, whereas about 59 percent of the sample judged the multiple-choice exam as being too tricky or tricky, only about 13 percent felt similarly about essay items, and only about 24 percent felt similarly about computerized type items. Furthermore, multiple-choice type exams were viewed as fair or very fair by about 55 percent, 37 percent for computerized type, and 14 percent for essay exams respectively.

The scale response distributions show that about 38 percent of the sample expected to receive high or very high scores on multiple-choice type exams, compared to only about 33 percent on computerize type exams and only about 21 percent on essay exams.

A meaningfully higher percentage of the sample (67 %) reported that essay exams are anxiety evoking, relative to only about 28 percent who felt similarly about multiple-choice exams, and only about 46 percent who felt similarly about computerized-type exams. Similarly, about twice the percentage of the prospective teachers (36 %) reported feeling ill at ease with computerized-type formats compared to multiple-choice type formats (20 %). Table 2 shows the sample means and standard deviations for the composite score and individual ratings of essay, multiple-choices exams versus computer –based exams.

Table 2.

Attitude scale Ratings of essay, multiple-choice type versus computer-based type exams for Means and Standard Deviations.

Scale	Essay		Multiple-choice		Computerized type	
	Mean	SD	Mean	SD	Mean	SD
Difficulty	2.16	0.92	2.82	0.79	2.48	0.63
Complexity	2.78	0.88	2.84	0.84	2.55	0.77
Clarity	2.84	0.86	3.09	0.77	2.94	0.78
Interest	2.30	1.00	2.80	0.91	3.22	0.71
Trickiness	3.75	0.78	2.44	0.99	3.29	0.65
Fairness	2.35	0.92	3.48	0.77	3.06	1.02
Value	3.02	0.87	3.61	0.54	3.43	0.78
Success	2.83	0.73	3.28	0.63	3.04	0.89
Anxiety	2.25	0.82	3.12	0.73	2.65	0.91
At ease	2.96	0.79	3.26	0.66	2.86	0.87
Average	2.72	0.86	3.07	0.76	2.95	0.80

Note: All the above scale ranged from 1 to 5 and was scored so that higher scores are indicative of more favorable test attitudes than lower scores.

On the whole, multiple-choice type exams (mean=3.07) were rated higher, on average, than essay (mean=2.72) type exams, and computerized type of exams (mean=2.95). Consistently higher mean ratings were observed for the multiple-choice type exam on 8 and 9 out of 10 individual scales appearing on the inventory, respectively for computerized type and essay type formats. Specifically, the multiple-choice type format was viewed as being easier than the essay type (3.26, 2.96 and 2.86), with about half (%51) of the prospective teachers judging multiple-choice exams to be very easy or easy in contrast to only about 12 % similarly perceiving the essay exam. Furthermore, multiple-choice exam was judged to be less complex (2.84, 2.78, and 2.55) and clearer (3.09, 2.84 and 2.94) than essay type exam.

In addition, prospective teachers tended to view the computerized type exam, in comparison to essay exam and multiple-choice exam, as relatively more interesting (3.22 and 2.80 and 2.30), less tricky (3.29, 3.75 and 2.44), and fairer(3.06, 2.35 and 3.48). With respect to the motivational variables assessed, the multiple-choice exam, in comparison to the essay exam and computerized type exam, was viewed as eliciting higher success expectancies (3.28 and 2.83 and 3.04), was perceived to be less anxiety evoking (3.12 and 2.25 and 2.65), and made respondents feel more at ease while taking the exam (3.26, 2.96 and 2.86).

Results devoted to relative rating scales

As mentioned, prospective teachers were also asked to directly compare and state their preference for one of the three item types with respect to a selected number of criteria, and provide reasons for their choices in each case. Following are some salient results, organized according to the major criteria for comparison among the formats.

Ease of preparation. The majority of sample (55 %) found it easier to prepare for essay exams than for multiple-choice (30 %) and computerized ones (15 %), because preparing for essay exams normally requires somewhat less time and effort for adequate preparation than preparing for multiple-choice type exams or computerized type-exams. Some of prospective teachers who found it easier to prepare essay exams also believed that the latter require a more profound mastery of the subjects matter material relative to other type exams. The minority of prospective teachers focused on different kinds of media that offer computerized-type exams that this makes easy to prepare.

Reflection of prospective teachers' knowledge. About 65 percent of the prospective teachers in the sample believed that grades on essay exams are more reflective indicators of the prospective teachers' knowledge of the exam material compared to grades on multiple-choice type exams (25 %) and computerized-type exams (10 %). The major reason offered is that essay exams provide prospective teachers the opportunity of accurately and optimally expressing their knowledge and ideas in writing. The remainder of the prospective teachers believed that multiple-choice exam scores are a more sensitive index of prospective teachers' knowledge, mainly because the latter normally cover a broader range of topics and a sample greater range of facts, concepts, and principles than typically is the case on essay exams. The minority of participants choose the computerized type exams because they believed that the more practice gives the more gain of knowledge.

Convenience of format usage. The majority (79 %) of the prospective teachers in the sample felt that the multiple-choice format is more convenient than the essay format, because there is no need to express answers in written form; it is possible to guess the correct answer with some probability of success; and a minimal amount of preparation is required for success. On the other hand, prospective teachers who found the essay format more convenient attribute this primarily to possibility of freely and accurately expressing ideas in writing(6 %), and 15 % felt that computerized-type format is more convenient, contemporary approach and easy to use.

Success expectancy. About 75 % of the prospective teachers in the sample believed that prospective teachers actually have a better chance of succeeding on multiple-choice relative to essay exams and computerized ones, for the following reason: multiple-choice exams, as a rule, are relatively easier than essay exams; the availability of options on multiple-choice type exams provide examinees with a sense of security and increased confidence while taking the test; examinee can guess (or copy) the correct answer; multiple choice exams preclude the possibility that examinees' scores will be unfairly lowered by grader on account of prospective teachers' spelling mistakes or poor writing abilities; and multiple-choice exams require less preparation and effort in order to succeed. The remainder of prospective teachers, who believed that they have higher probability of succeeding on essay exams, attributed high expectancies mainly the fact that essay exams allow prospective teachers, in principles, to give expression to their maximum degree of knowledge on the given subjects. They further believed that tendency for teachers' subjective grading of essay papers works to the advantage of prospective teachers, thus increasing their grades and probability of success on the exam.

Perceived fairness. To about more than half (55%) of the sample, multiple-choice exams were perceived to be more fair than essay and computerized type exams, for two main reasons: the nil probability of guessing the correct answer assures the examinees' scores reflect actual knowledge rather than luck of error and also prospective teachers are offered the possibility of accurately expressing and elaborating on ideas in essay exams, but the essay exams is more

subjective. The remainder of the sample believed that multiple-choice type exams are fairer than essay exams mainly because of the partial information provided prospective teachers by the availability of options, and the freedom from having to construct and present the answer in written form.

Degree of anxiety evoked. The vast majority (67 % and 46 %) of the prospective teachers reported that taking an essay and computerized exam (respectively) are more anxiety evoking than taking a multiple-choice exam because additional effort is expended and emotional energy is demanded of prospective teachers having to select, organize, and express ideas in essay form. Further, there is a total absence of information or clues leading to the correct answer as well as marked degree of over-learning required to succeed on essay exams; relatively greater length and complexity of responses are required in construction type items. The minority of prospective teachers who reported that multiple-choice type exams are relatively more anxiety evoking attribute this mainly to the difficulty and stress involved in choosing among given options, the relatively large number of items prospective teachers normally have to respond to on multiple-choice exams, and the increased probability of error.

Overall preference. If had to chose one, about three quarters of the sample (70 %) clearly reported an overall preference for multiple-choice over essay exams, for four main reasons: a) the availability of options to choose from, b) the convenient item format, c) the freedom from having to organize and write the answer, and d) the possibility of guessing or copying the correct answer. The minority (24%) of the prospective teachers reported a preference for essay over multiple-choice type exams, attributing their choice mainly to a) the possibility of accurately communicating ideas in written form, b) simplicity of the item format, and c) the possibility of obtaining some credit for a partially correct response. The majority of sample (65%) indicated that they are willing to use combination of two exam types (essay and multiple-choice) in their examinations.

In the last question of the instrument, I asked if they knew alternative types of assessment and will use in their further teaching. The majority of them (92 %) gave the names of some new type of assessment (like, portfolios, simulations, case-based evaluation, presentations, etc) however they said they are willing and/or try to use the new type of assessment only 42 % of them. 65 % said that they will definitely use combination of (essay and multiple-choice type) two exams.

In summary, the data presented here pointed to a more positive attitudinal disposition of prospective teachers towards multiple-choice relative to essay type exams with respect to the majority of dimensions assessed; however, they do not want to choose one to another, because they are willing to use many assessment types altogether or combination of at least two of them.

Discussion and Conclusions

The data presented in this study indicated that multiple-choice type of exams are generally perceived more favorably than essay type and computerized type items along most dimension assessed. Many favored of multiple-choice type test, along the dimensions of perceived difficulty, anxiety, success expectancy, complexity, and feeling at ease with the format. Zeidner (1987) and Traub and McRury (1990) found similar results. Nevertheless, Birenbaum and Feldman (1998) discovered on one hand that students with good learning skills, who have high confidence in their academic ability, tend to prefer the essay type of

assessment to the multiple-choices of examinations. Conversely, students with poor learning skills, who tend to have low confidence in their academic ability, prefer the choice over the constructed-response type of assessment. Also smallest differences between the formats were evidenced on the dimensions of trickiness, perceived interest, and perceived value in favor of computerized-type exams. Many see the computerized type exam is more contemporary approach and also many have a positive attitude to use it in their further teaching, but somehow they are not comfortable and convenient to use yet.

The data clearly indicated that perspective teachers perceived multiple-choice items more favorably than essay type items. It is not surprising that over the past two decades, in Turkey the multiple-choice type tests have been using in many areas of selecting and evaluating purposes from high schools to universities. Multiple-choice formats, or an emphasis on detailed factual answers, push students towards a surface approach, while open, essay-type questions tend to encourage a deep approach (Entwistle & Entwistle, 1991). This result becomes reinforced by the finding that a change from a multiple-choice to essay-type examinations had shifted the overall tendency of the students from a surface approach towards a deep approach (Thomas & Bain, 1984). However, Students' perception about evaluation and assessment in higher education the reverse relationship between assessment and the student's approach to learning is evidenced. Entwistle and Tait (1990) found that students who reported themselves as adopting surface approaches to learning preferred teaching and assessment procedures which supported that approach, whereas students reporting deep approaches preferred courses which were intellectually challenging and assessment procedures which allowed them to demonstrate their understanding (Entwistle & Tait, 1995).

Sambell and McDowell (1998) and Sambell et al. (1997) tried to unveil students' interpretations, perceptions and behaviors when experiencing different forms of alternative assessment and more in particular its consequential validity or the effects of assessment on learning and teaching. Although most assessment formats are perceived to be fairer than their conventional partners, there were some concerns about the reliability of self and peer assessment, even though students valued the activity. Sambell et al. (1997) stated that from the student perspective the issue of fairness is important, and includes more than only the possibility of cheating. In this respect, students criticize the more conventional evaluation methods. The use of test attitude inventories on large-scale and routine basis in the classroom might serve to fill the needed gap for a judgment approach to the face validity of the classroom test and their constituent components, providing instructors and educational researchers with useful information about key dimensions in the test situations.

It is plausible that prospective teachers' attitudes and perceptions about the test form is important factors in affecting their test preparation behavior; their cooperation and test motivation during the exam; and possibly influencing the level of their test performance and attainment on the exam. The strong preference of students examines for multiple-choice over computerized and essay type formats evidenced in this research deserved to be given due weight and careful consideration by educational specialists. A study that investigated the influence of assessment type and discipline of study on students' learning approaches indicated that assessment type had no significant influence on how students approach their learning, while discipline did have significant influence on student learning (Watkins, 1982; Ramsden, 1988; Scoullar, 1998; Smith and Miller, 2005). Furthermore, prospective teachers' attitudes and perceptions with respect to test forms are important pieces of information for the instructors and measurement specialist alike, since they serve as indicators of a test's face validity from the point of view of the most affected by the test results. As point out in the

literature, the concept of face validity implies that a test should not only valid from a content, construct, or perspective validity point of view, but also appear to be valid to a variety of judges-including test takers (Nevo, 1985).

Within the research on students' perceptions about alternative assessment, contradictive results are found. For example, although it seemed that peer and tutor scores correlated with each other, Orsmond et al. (1997) revealed little agreement between student marks and between the student's mark and the tutor scores, with poor students tending to over-mark their work, whilst good students tended to under-mark. Although much disagreement was found, students valued this self-assessing (and evaluating others) exercise. They thought that self-assessment made them think more critical and students felt that they learned more and worked in a more structured way. Mires et al. (2001) found significant correlations between student's scores and the tutor score, but students failed to acknowledge the values of self-assessment in terms of feedback and as a learning opportunity, and expressed uncertainty over their marks. Students perceived many more disadvantages (including being more stressful, uncertainty about capability, not knowing how to mark, anxiety about failure, being accused of cheating or marking too low) than advantages (for example seeing mistakes) in the self-marking exercise (Mires et al., 2001). Challis' (2001) commented that each assessment method simply needs to be seen in terms that recognize its own strengths and its differences from other methods, rather than as a replacement of any other assessment methods and procedures.

Smith and Miller (2005) indicated that the context and the assessment mode make the student's approach to learning a very individual approach that changes constantly. In this manner, students' perceptions of assessment become very arbitrary and their value for educational practices should be called in question. However, most research data show patterns, tendencies, and relations between students' perceptions, the different assessment methods and student learning that provide useful insights for student educators, though the web of influence is yet far from clear (Smith and Miller, 2005).

Although the present research has produced some interesting and potentially useful findings, there are limitations in methodology and research design. Further research to extend and explore current findings might include the following:

- It should be held in mind that the study was conducted among volunteer prospective teachers from four different departments (of physics education, science education, computer education, and elementary teacher education) only. It might very well be that different results would have been obtained for other age groups or students in different educational or cultural settings. Therefore, future study is needed in order to extend the validity of the finding beyond these groups studied and the specific educational and cultural settings.
- Investigating students from other universities to enhance the generalizability of present findings beyond the instructional and assessment policies and practices of one institution, and the student selection policies of a single university.
- Requiring students to report on their actual learning behaviors near the conclusion of their preparation for a particular type of examination, or immediately after completing their examinations. This will give a clearer picture of how different assessment types might

influence student learning, and will serve to remove the hypothetical conditions feature of the research reported in this paper.

References

- Bartram, D. (2005). Computer-based testing and the Internet. In A. Evers, O. Smit-Voskuyl, & N. Anderson (Eds.). *The Handbook of Selection*. Oxford: Blackwell.
- Bartram, D & Hambleton, R. K. (2006). *Computer-Based Testing and the Internet*. John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester.UK.
- Biggs, J. B. (1993). What do inventories of students' learning processes really measure? A theoretical review and clarification. *British Journal of Educational Psychology*, 63, 3–19.
- Birenbaum, M. & Feldman, R. A. (1998) Relationships between learning patterns and attitudes towards two assessments formats, *Educational Research*, 40(1), 90–97.
- Claxton, C., & Murrell, P. (1987). Learning styles: Implications for improving education practices. ERIC Document Retrieval Service No ED293478.
- Carless, D., Joughin, G, & Mok, M.M.C. (2007). Learning-oriented assessment: principles and practice. *Assessment & Evaluation in Higher Education*, 31(4), 395–398.
- Challis, M. (2001) Portfolios and assessment: meeting the challenge, *Medical Teacher*, 23(5), 437–440.
- Entwistle, N. J. (1996). Learning and studying: Contrasts and influences. Seattle: New Horizons for Learning. Retrieved March 20, 1997 from <http://www.newhorizons.org/crfutentwistle.html>
- Entwistle, N. J. & Entwistle, A. (1997) Revision and experience of understanding, in: F. Marton, D. Hounsell & N. Entwistle (Eds) *The experience of learning. Implications for teaching and studying in higher education* (Edinburgh, Scottish Academic Press), 146–158.
- Entwistle, N. J. & Tait, H. (1990) Approaches to learning, evaluations of teaching, and preferences for contrasting academic environments, *Higher Education*, 19, 169–194.
- Entwistle, N. & Tait, H. (1995) Approaches to studying and perceptions of the learning environment across disciplines, *New Directions for Teaching and Learning*, 64, 93–103.
- Gronlund, N. E. (1976). *Measurement and evaluation in teaching*. 3rd ed. New York: Macmillan.
- Hambleton, R. K., & Pitoniak, M. J. (2002). Testing and measurement: Advances in item response theory and selected testing practices. In J. Wixted (Ed.), *Stevens' handbook of experimental psychology* (3rd ed., Vol. 4, pp. 517–561). New York: Wiley.

- Hibberd, S. (1996). The mathematical assessment of students entering university engineering courses, *Studies in Educational Evaluation*, 22(4), 375-384.
- Irvine, S. H., & Kyllonen, P. C. (Eds.). (2002). *Item generation for test development*. Hillsdale, NJ: Erlbaum.
- Luecht, R. M. (1998). Computer-assisted test assembly using optimization heuristics. *Applied Psychological Measurement*, 22, 224–236.
- Luecht, R. M., & Clauser, B. E. (2002). Test models for complex computer-based testing. In C. N. Mills, M. T. Potenza, J. J. Fremer, & W. C. Ward (Eds.), *Computer-based testing: Building the foundation for future assessments* (pp. 67–88). Hillsdale, NJ: Erlbaum.
- Mead, A. D., & Drasgow, F. (1993). Equivalence of computerized and paper-and-pencil cognitive ability tests: A meta-analysis. *Psychological Bulletin*, 114, 449–458.
- Merrell, C., & Tymms, P. (2007). Identifying Reading Problems with Computer-Adaptive Assessments. *Journal of Computer Assisted Learning*, 23(1), 27-35.
- Mires, G. J., Friedman Ben-David, M., Preece, P. E. & Smith, B. (2001) Educational benefits of student self-marking of short-answer questions, *Medical Teacher*, 23(5), 462–466.
- Nevo, B. (1985). Face validity revisited. *Journal of Educational Measurement*, 22, 287-293.
- Nevo, B. (1986). The practical value of examinee's feedback questionnaires. Paper presented at the 21st Congress of International Association of Applied Psychology, Jerusalem, Israel.
- New Horizons for Learning. (2002). Assessment Terminology: A Glossary of Useful Terms. [URL:<http://www.newhorizons.org/strategies/assess/terminology.htm>] (Retrieved on Feb. 15, 2007).
- Olsen, J. B., Maynes, D. D., Slawson D., & Ho, K. (1986). Comparison and equating of paper-administered, computer-administered and computerized adaptive tests of achievement. Presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA,
- Pitoniak, M. (2002). Automatic item generation methodology in theory and practice (Center for Educational Assessment Research Report No. 444). Amherst, MA: University of Massachusetts, Center for Educational Assessment.
- Ramsden, P. (1988). Context and strategy: Situational influences on learning. In R. R. Schmeck (Ed.), *Learning strategies and learning styles* (pp. 159–184). New York: Plenum Press.
- Sambell, K., McDowell, L. & Brown, S. (1997). But is it fair?: an exploratory study of student perceptions of the consequential validity of assessment, *Studies in Educational Evaluation*, 23(4), 349–371.

- Sambell, K. & McDowell, L. (1998) The construction of the hidden curriculum: messages and meanings in the assessment of student learning, *Assessment and Evaluation in Higher Education*, 23(4), 391–402.
- Scoullar, K. (1998). The influence of assessment method on students' learning approaches: Multiple choice question examination versus assignment essay. *Higher Education*, 35, 453–472.
- Scoullar, K. (2000). The influence of assessment method on students' learning approaches: Multiple choice question examination versus assignment essay. *Studies in Higher Education*, 25(3), 309 – 323.
- Smith, S. N. & Miller, R. J. (2005). Learning approaches: examination type, discipline of study, and gender. *Educational Psychology*, 25(1), 43 – 53.
- Thomas, P. R. & Bain, J. D. (1984) Contextual dependence of learning approaches: the effects of assessment, *Human Learning*, 3, 227–240.
- Thorndike, R.L. (1982). *Applied Psychometrics*. Boston: Houghton-Mifflin.
- Thorndike, R.L., & Hagen, E. (1969). *Measurement and evaluation in psychology and education*. 3rd ed. New York.
- Traub, R. E. & MacRury, K. (1990) Multiple-choice vs. free response in the testing of scholastic achievement, in: K. Ingenkamp & R. S. Jager (Eds) *Test und tends 8: jahrbuch der pädagogischen diagnostik* (Weinheim und Base, Beltz Verlag), 128–159.
- van der Linden, W. J., & Glas, C. A. W. (Eds.). (2000). *Computer adaptive testing: Theory and practice*. Boston, MA: Kluwer.
- Wainer, H., et al. (Eds.). (2000). *Computerized adaptive testing: A primer* (2nd ed.). Hillsdale, NJ: Erlbaum
- Watkins, D. (1982). Identifying the study process dimensions of Australian university students. *The Australian Journal of Education*, 26, 76–85.
- Zeidner, M. (1985). *Psychological testing: The examiner's perspective*. Symposium conducted at the annual meeting of the Israeli Psychological Association, Ramat-Gan, Israel.
- Zeidner, M. (1986). Situational bias-the examinee's perspective. In B. Nevo and R.S. Jager (Eds.), *Psychological testing: The examinee perspective*. Göttingen: Hogrefe Intl.
- Zeidner, M. (1987). Essay versus multiple-choice type classroom exams: The students' perspective. *Journal of Educational Research*, 80(6), 352-358.
- Zenisky, A. L., & Sireci, S. G. (2002). Technological innovations in large-scale assessment. *Applied Measurement in Education*, 15(4), 337–362.