

Appraising Pre-service EFL Teachers' Assessment in Language Testing Course Using Revised Bloom's Taxonomy

Elham Mohammadi

Department of ELT, Tarbiat Modares University, Tehran, Iran

E-mail: e.mohammadi84@gmail.com

Gholam Reza Kiany (Corresponding author)

Department of ELT, Tarbiat Modares University, Tehran, Iran

E-mail: rezakiany_gh@modares.ac.ir

Reza Ghaffar Samar

Department of ELT, Tarbiat Modares University, Tehran, Iran

Ramin Akbari

Department of ELT, Tarbiat Modares University, Tehran, Iran

Received: 27-10- 2014

Accepted: 19-01- 2015

Advance Access Published: February 2015

Published: 01-07- 2015

doi:10.7575/aiac.ijalel.v.4n.4p.8

URL: <http://dx.doi.org/10.7575/aiac.ijalel.v.4n.4p.8>

Abstract

The teachers need to be conceived as a “change agent” and not as a mere transmitter of knowledge and culture. In developing countries like Iran, one of the most significant concerns in the field of teachers’ education is efficiency of pre-service programs. To this aim, the current descriptive-evaluative study intended to describe the state of pre-service teachers' assessment in the field of language testing by (a) examining the exam questions to find out whether they are aligned with curriculum objectives and syllabus (content validity, (b) exploring whether they take care of higher order cognitive processes and (c) finding what combinations of cognitive process levels and knowledge types in Revised Bloom's Taxonomy are prevalent in the questions. The results exhibited an unbalanced coverage of content in exams. Also the questions were found to be inadequate in terms of measuring complex cognitive skills (Analyze and Evaluate); Remember and Understand domains take up 91.6 % of all questions and no item was found for Create. Three combinations of cognitive process level and knowledge type was dominant in the data set: (1) Remember Factual Knowledge, (2) Understand Conceptual Knowledge, and (3) Apply Procedural Knowledge. These associations confirm the Anderson and Krathwohl's (2001) proposition.

Keywords: pre-service teachers, assessment, assessment literacy, Revised Bloom's Taxonomy, cognitive processes, knowledge types

1. Introduction

1.1 Statement of the Problem

Rationale of curriculum planning and different reforms in the domain of education system in Iran and world has consistently indicated improvement of assessment-measurement system and its alignment with theories and ultimate goals of education as being the most salient element of success in educational reform and fundamental shift in quality of school teaching and learning (Rabihavi, et al. 2011). Assessment determines whether the process of learning leads to success or failure (Dochy, 2009; Kozhageldiyeva, 2005). To find out whether evaluation process has been carried out accurately needs evaluation itself.

Popham (1987) has referred to measurement-oriented reforms in education and believes amending educational movement results in curriculum reform. After examining the role played by assessment in reforming science curriculum, Orpwood (2001) also found out lack of proper attention to assessment as a reason for some of the problems. He reminds the demand on experts for more attention towards the role of assessment.

In recent years the term 'assessment literacy' has been coined to denote what teachers need to know about assessment (Huttner, et al., 2011). Traditionally it was regarded as the ability to select, design and evaluate tests and assessment procedures as well as to score and grade them on the basis of theoretical knowledge. More recent approaches embrace a broader understanding of the concept when taking the implications of assessment for teaching into account. Two key questions are asked: (a) What does an assessment tell students about the achievement outcomes we value? (b) What is likely to be the effect of this assessment on students? (Stiggins, 1991). It's essential for assessment literates to know and understand the key principles of sound assessment and translate these into quality information about students' achievements and effective instruction. Boyles (2005, p. 18) points to language teachers' need of the necessary tools for

analyzing and reflecting upon test data in order to make informed decisions about instructional practice and program design. In their views (Boyles, 2005; Stiggins, 1991) the notion of assessment literacy extends beyond technical knowledge about how to select and create appropriate assessment instruments for specific purposes to include the ability to analyze empirical data to improve instruction. In other words, being literate in assessment involves a move away from a passive interpretation towards an active application of data that will impact on teaching and assessment is perceived as a means to promote learning rather than merely observe and record it, hence 'assessment' for learning (Stiggins, 2002).

Several studies in a range of educational contexts indicates that "the typical teacher can spend as much as a third to a half of his or her professional time involved in assessment-related activities" (Newsfields, 2006; Stiggins, 1999 p. 23). Additionally in the global economy of the 21st century, students will need to understand the basics, but also to think critically, to analyze, and to make inferences. Helping students develop these skills will require changes in assessment at the school and classroom level, as well as new approaches to large-scale, high-stakes assessment. These points imply the need for future teachers to recognize the importance of higher-order thinking and problem solving skills. They are to be their students' model of learning that makes the mission of teacher preparatory programs and teacher trainers more demanding. These programs are to create essential sensitivity among teacher trainees toward assessment. This study seeks to investigate the state of teacher trainees' assessment in the field of language testing to identify potential weaknesses and inconsistencies.

1.2 Significance of the Study

Iranian education system is faced with several complications (Rabihavi, et al. 2011; Molaenezhad & Zekavati, 2007); among these, evaluation practice can be mentioned which entirely relies on retrieving knowledge. It suffices for learners to memorize content and use only retrieving capacity. On many occasions this evaluation system fails to assess and measure abilities like reasoning, processing, synthesizing and accurate judgment among learners. The questions constructed and administered by teachers are necessary instruments for measuring such abilities. Examining teachers' exam questions at school has found these questions and progress tests incompatible with measurement criteria (Rabihavi, et al. 2011). This might be attributed to the state of teacher education in Teacher Training Centers; Teacher educators were not able to create due attention to testing standards and psychometric principles among student teachers.

For making sound inferences about students' abilities and subsequently directing their teaching, language teachers need a strong knowledge of assessment practices. Only if do they understand the basic principles of classroom assessment their efforts to improve teaching and learning based on assessment results will be effective. Addressing relevant questions, findings of the current study would help understand the problem better and can be used for improving and reforming evaluation system and teaching methods. They may somehow shed light on needed curriculum in teacher training domain. Given the significance of assessment courses in general and language assessment literacy in particular, there is a need to take a closer look at the status quo of these courses in Iranian context.

1.3 Foreign Language Testing

Brindley (2001) has referred to the allegedly heated issue of future language teachers' preparation in the field of foreign language testing and evaluation in recent years. O'Loughlin (2006, p. 71) argued that second language assessment is a "notoriously difficult domain of knowledge" for students in second language teacher education programs because of the high theoretical complexity of key concepts like reliability, validity, and practicality and the need to be balanced against each other in designing and using assessment instruments. Developments in the domain of language testing are different from language teaching; however they are closely tied to each other (Johnson & Johnson, 2001, p. 187). Construction of good test items is believed to be a demanding task for teachers since it involves "a psychological understanding of pupils, sound judgment, persistence, and a touch of creativity" as well as field knowledge and clear view of the desired outcomes (Groulund, 1985).

Jin (2010, p. 556) brings up the urgent need for teachers to be thoroughly trained in language assessment concepts, skills and strategies. Assessment is an increasingly important domain of language teachers' expertise as the professional demands on them to accurately assess their students increases as the theory and practice of assessment continues to mature (O'Loughlin, 2006, p. 71; Newsfields, 2006; Brindley, 2001). Shohamy (2005, p. 107) also argues that professional development in assessment is not a question merely of demonstrating the technical 'tricks of the trade'. She argues for "... the need to expand the role of teacher education programs in which teachers are exposed not only to the procedures and methods of testing and assessment but also to aspects related to the consequences of tests." Very recently Lam (2014) explored the overall language assessment training landscape in primary/secondary school contexts and investigated the extent to which two assessments courses may facilitate and/or inhibit the development of pre-service teachers' language assessment literacy in one teacher education institution. His findings reveal that language assessment training in Hong Kong remains inadequate and the courses are still unable to bridge the theory-practice gap within the assessment reform context.

Very little research as yet has been undertaken to study the students' evaluation of language testing and assessment courses (Inbar Lourie, 2008; Kleinsasser, 2005; O'Loughlin, 2006). Studies conducted by Bailey and Brown (1996), Brown and Bailey (2007), and Jin (2010) are among the few ones on language testing courses and included large groups of informants teaching language testing courses. Jin (2010) used almost the same questionnaire utilized by Bailey and Brown (1996) and found out most of the topics covered are the same as the ones listed by Brown and Bailey in their studies. He concluded that in China LTA (i.e Language Testing and Assessment) courses cover essential theoretical and

practical aspects of the language testing area. In two other studies by O'Loughlin (2006) and Kleinsasser (2005) more comprehensive reports on language assessment courses were offered. O'Loughlin focused on a post-graduate elective course titled 'Assessment in the language classroom'. The course entails practical components (e.g. designing assessment tools for assessing various skills) as well as conceptual themes. The author attempted to reveal how students' understanding of key concepts and ability to evaluate current assessment documents and instruments develops. The findings indicate that participants (two students) attained the course objectives; yet they differed in their readiness and capacity to grasp new ideas in language assessment area. The researcher attributed this difference to factors such as students' cultural background, prior experience with assessment as learners and teachers and the characteristics of the input they receive in language assessment classes. Hence he concluded that a learner-centered approach should be adopted which takes those factors into account while planning and managing assessment courses. Kleinsasser (2005) like O'Loughlin (2006) presented readers with direct and detailed reports related to language assessment courses and showed the significance of the collaboration between teachers and students for improving the quality and usefulness of an LTA course.

1.4 Theoretical Framework

To get a better in-depth understanding of the assessment practices of teacher preparatory programs, a robust cognitive framework is required. Revised Bloom's taxonomy has been extensively applied in testing and evaluation across different disciplines and its efficiency is substantiated time and again (Chen, 2004; Squire, 2001; Aviles, 1999). Bloom's taxonomy is a framework for examining the depth of cognitive process levels in educational objectives and it is used to determine the extent to which assessment tools measure higher-order thinking skills. Revised Bloom's taxonomy provides a broader vision of learning that includes not only acquiring knowledge but also being able to use knowledge in a variety of new situations. This taxonomy, a revision of the original Bloom's taxonomy, was developed by Anderson & Krathwohl (2001). It has two dimensions of cognitive process and knowledge which is extensively explained in Appendix A. Cognitive dimension includes six major categories and 19 specific sub-categories whereas knowledge dimension contains four main categories. Although relatively few studies in Iranian teacher education context used this taxonomy for examining assessment tools (Rabihavi, et al. 2011), numerous studies worldwide adopt it to take a closer look at evaluation procedures. Masters et al. (2001) investigated randomly-selected multiple-choice questions from 17 test-banks of accompanying selected nursing textbooks. 2143 items were rated on thirty generally accepted guidelines for writing multiple-choice questions, the cognitive levels of original Bloom's taxonomy and distribution of correct answers. Results indicated that most of the questions (i.e. 47.3 %) were written at the lowest cognitive level of Knowledge. 24.8 % and 21 % of items were respectively at Comprehension and Application levels and only 6.5 % were at the Analysis level. Since most of the textbooks were intended for upper division courses, these findings turn out to be surprising. Another field of study whose tests were the subject of scrutiny with regard to cognitive process levels is Agriculture (Squire, 2001). 628 agricultural science questions in senior secondary schools in Botswana were analyzed. A great proportion of questions were at Knowledge level of Bloom's original taxonomy. Almost no item was found at higher cognitive levels of Application, Analysis, Synthesis, and Evaluation. Unexpectedly even the essay-type items in those tests were at the lowest levels. Using Revised Taxonomy, Chen (2004) examined the knowledge types and cognitive levels of Computer Science test in Technical College Entrance Examination of Taiwan between 2001 and 2004. In a similar pattern, most items (44 % to 77 %) assess only lower-level thinking that required students to remember factual information. No item was found at Evaluate and Create levels.

1.5 Research Questions

Given the above introduction the research questions are formulated in the following way:

Q1: Do exam questions of Teacher Training centers cover all cognitive levels of Bloom's Taxonomy properly and consistent with objectives of the curriculum?

Q2: Do exam questions of Teacher Training centers have content validity?

Q3: What cognitive process levels and knowledge types in the Revised Bloom's Taxonomy were measured in Testing exams of Teacher Training Centers?

2. Method

Concerning the nature and purpose of the study, we adopt a descriptive-evaluative research method. To this aim the curriculum, syllabus and the assessment instrument for language testing course of pre-service teachers are examined and analyzed.

2.1 Context of the Study

The main bodies responsible for training teachers in Iran are Teacher Training Centers and Universities that are jointly managed by Ministries of Education, and Science, Research, and Technology. Prior to 2013, the undergraduate teacher training programs consisted of two two-year period; upon completion of the first two-year the graduates would get an Associate's degree in relevant majors. The then-teachers are allowed to continue their studies for another two-year period to obtain a B.A. This second term was mostly for in-service teachers. In 2013 the teacher training programs shifted onto four-year long training at the end of which graduates would receive a B.A.

2.2 Sampling Procedures

Five Teacher Training Centers were chosen for this study. Two reasons can be enumerated for this selection:

- (1) The Centers' ease of accessibility to the researcher (Convenience Sampling)
- (2) To the extent the availability is possible for the researcher, it was tried to cover as geographically different areas as possible (Purposive Sampling)

The centers are:

- (a) *Shahid Bahonar Center (located in Tehran)*
- (b) *Shahid Mofatteh Center (located in Shahr-E-Rey)*
- (c) *Shahid Bahonar Center (located in Arak)*
- (d) *Mirshaki Center (located in Aligoudarz, Boroujerd)*
- (e) *Shahid Rajaei Center (located in Urmia)*

Language Education Departments in T. T. Centers follow a curriculum prescribed by High Council of Planning under the supervision of Ministry of Science, Research and Technology and Coordinating Body of Teacher Education in 2001. Short descriptions for each of the core courses is provided in this document, along with educational objectives, syllabus (the weight given to each section or topic), and suggested materials. However lecturers are allowed to structure their own lessons by keeping those points in mind. Language testing (language assessment literacy) is divided into two two-credit courses titled Testing (1) and Testing (2). The assessment instruments were designed and developed by the lecturer him/herself. In recent decades teacher-made tests have played a very fundamental role in assessing students' learnings, however, their efficiency has rarely been investigated (Rabihavi, et al., 2011).

2.3 Data Analysis

The exam questions for Testing 1 and 2 were collected, and then each question was rated based on Revised Taxonomy's cognitive processes and knowledge types (explained below). To ensure the accuracy of researchers' coding results two Ph.D graduates of Tarbiat Modares University in TEFL were recruited to serve as raters. The criteria for coding, the definition of the Revised Bloom's taxonomy (Anderson & Krathwohl, 2001) and some sample questions were given to the raters to help them become familiar with the coding framework. Then Cohen's kappa coefficient was computed to ensure inter-rater reliability and to observe whether there is a statistically significant consistency among raters.

To see the frequency counts and percentages of the major combinations as well as the sub-combinations of the cognitive process and knowledge dimensions crosstabulation which is a descriptive statistical procedure was employed.

3. Results

The consistency among raters was specified by computing kappa value. This figure was respectively 0.88 and 0.83 for Rater 1 * Rater 2 and Rater 1 * Rater 3 that shows a high degree of agreement among raters. Both values are significant at .001 level. This means

The results for the first research question are reported below.

For answering research questions Table of Specifications is needed. A TOS, sometimes called a test blueprint, is a table that helps teachers align objectives, instruction, and assessment (Notar, Zuelke, Wilson, & Yunker, 2004). When constructing a test, teachers need to be concerned that the test measures an adequate sampling of the class content at the cognitive level that the material was taught. The TOS can help teachers map the amount of class time spent on each objective with the cognitive level at which each objective was taught thereby helping teachers to identify the types of items they need to include on their tests. The tables developed for this study are provided in the Appendix B.

$$\text{Importance percentage of each chapter} = \frac{\text{No. of sessions devoted to each chapter}}{\text{No. of all sessions}} * 100$$

No. of questions for each chapter = *no. of all questions * importance percentage of each chapter*

Concerning the question on cognitive domain of exam questions, the results (overall frequency and percentage) for all five centers are reported in the following tables. The purpose of this question was to determine the extent to which exam questions are successful in measuring higher-level thinking processes (e.g. analyze, evaluate, and create).

The questions were judged on Bloom's Revised Taxonomy. Overall 346 items out of 441 fall within the lowest level of cognitive process that would mean as 78.5% in percentile. This category entails two subcategories of recognizing and recalling which respectively constitute 38.3 % and 40.1% of this category.

Concerning next cognitive level, Understanding, 58 items were found which would mean 13.2%. Among seven different subcategories, Interpreting comprises 36.2% of the whole category. Other processes of Exemplifying, Classifying, Inferring, Comparing and Explaining make up respectively 1.1%, 1.1%, 2.7%, 1.6%, 1.8% of all questions. No item was found for Summarizing.

Table 1 demonstrates the distribution of items across sub-categories of Understanding.

Table 1. Cognitive level (Understand)

		Frequency	Percent	Valid percent	Cumulative percent
Valid	Interpreting	21	4.8	36.2	36.2
	Exemplifying	5	1.1	8.6	44.8
	Classifying	5	1.1	8.6	53.4
	Inferring	12	2.7	20.7	74.1
	Comparing	7	1.6	12.1	86.2
	Explaining	8	1.8	13.8	100.0
	Total	58	13.2	100.0	
Missing	System	383	86.8		
	Total	441	100.0		

Cognitive process of apply which consists of executing and implementing processes constitutes 6.8% of questions (30 items out of 441). Executing comprises 56.7% and Implementing 43.3% of the main category. Given next cognitive process, Analyze, no question was found on subcategories of Differentiating and Attributing; Organizing makes up 0.7% of all items. Evaluate is divided into Checking and Critiquing; each one of these subcategories makes 0.5% of all questions that means two items for each. No item was found for the highest cognitive level, i.e Create, in the taxonomy.

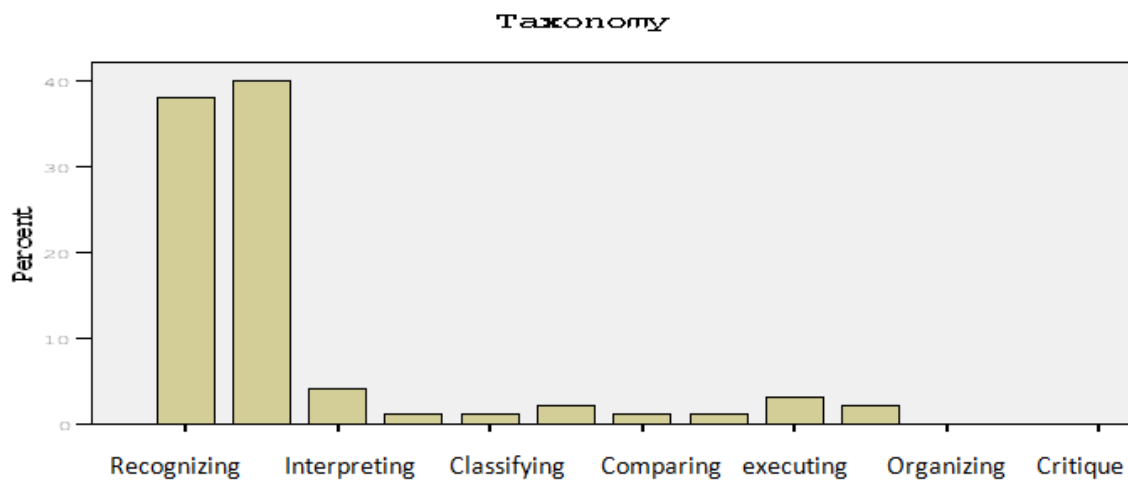


Figure 1. The percentage of questions on cognitive categories (Revised Bloom's Taxonomy)

Figure 1 displays the resulting item analysis in terms of coverage of cognitive sub-categories. There is a huge difference between the first major category (i.e Remember), consisting of two sub-levels of Recognizing and Recalling, and the rest of the taxonomy.

Table 2. The frequency and percentage of exam questions on cognitive subcategories

		Frequency	Percent	Valid percent	Cumulative percent
Valid	Recognizing	169	38.3	38.3	38.3
	Recalling	177	40.1	40.1	78.5
	Interpreting	21	4.8	4.8	83.2
	Exemplifying	5	1.1	1.1	84.4
	Classifying	5	1.1	1.1	85.5
	Inferring	12	2.7	2.7	88.2
	Comparing	7	1.6	1.6	89.8
	Explaining	8	1.8	1.8	91.6
	Executing	17	3.9	3.9	95.5
	Implementing	13	2.9	2.9	98.4
	Organizing	3	.7	.7	99.1
	Checking	2	.5	.5	99.5
	Critiquing	2	.5	.5	100.0
	Total	441	100.0	100.0	

For answering the second question which is on the content validity, table of specification, objectives and topics set in the curriculum and syllabus are taken into account (Appendix B). In the following table the coverage of subject matters in exam questions is specified in terms of the number of items and the percentage for Testing (1):

Table 3. The amount of content coverage in Testing (1) exams regardless of teacher training center

Content	Number of items	Percentage of items
Preliminaries	37	16.1 %
Functions of Language Tests	32	13.9 %
Forms of Language Tests	22	9.5 %
Basic Statistics in Testing	52	22.6 %
Test Construction	29	12.7 %
Characteristics of a Good test	46	20 %
Theories of Language Testing	12	5.2 %
Total	230	100 %

For Testing (1), Table 3 shows the relevant proportion of items on each subject matter. Looking closely it is apparent the majority of questions (62%) are on purely theoretical areas which are taught in the first half of the semester. The sections on "Test Construction", "Characteristics of Good Test", and "Theories of Language Testing" are not satisfactorily taken care of. Given their importance that derives from the practical value essential for future teachers, it is necessary to amend this weakness and devote a larger number of questions to these areas.

Table 4. The amount of content coverage in Testing (1) exams across teacher training centers

Content	Urmia center		Boroujerd center		Arak center		Shahr-e-Rey center		Tehran centers		Total
Preliminaries	8	20.5 %	4	9.7 %	10	18.9 %	9	14.5 %	6	17.1 %	37
Functions of Language Tests	8	20.5 %	7	17 %	5	9.4 %	8	13 %	4	11.4 %	32
Forms of Language Tests	3	7.7 %	7	17 %	9	17 %	2	3.2 %	1	2.8 %	22
Basic Statistics in Testing	10	25.6 %	8	19.5 %	11	20.7 %	15	24.2 %	8	22.8 %	52
Test Construction	3	7.7 %	4	9.8 %	5	9.4 %	11	17.7 %	6	17.1 %	29
Characteristics of a Good test	6	15.4 %	9	22 %	8	15.1 %	14	22.6 %	9	25.7 %	46
Theories of Language Testing	1	2.6 %	2	4.9 %	5	9.4 %	3	4.8 %	1	2.8 %	12
Total	39	100%	41	100 %	53	100 %	62	100 %	35	100 %	230

Examining the centers individually (Table 4), all of them lack enough coverage of the last three sections (mentioned above). This, in effect, exhibits a major drawback of teacher preparatory programs which is their theoretical orientation. Shahid Mofatteh and Shahid Bahonar (Tehran) Testing exams include comparatively sufficient number of questions on the topics of the syllabus in comparison with other centers.

Table 5. The amount of content coverage in Testing (2) exams regardless of teacher training center

Content	Number of items	Percentage of items
Testing Vocabulary	26	12.3 %
Testing Structure	28	13.3 %
Testing Pronunciation	25	11.8%
Testing Listening Comprehension	27	12.8 %
Testing Oral Production	29	13.7 %
Testing Reading Comprehension	22	10.4 %
Testing Writing Ability	19	9 %
Cloze and Dictation Type Tests	20	9.4%
Functional Testing	15	7.1 %
Total	211	100%

By taking a look at Table 5 that includes all questions of Testing (2), a rather balanced and satisfactory coverage of content areas is observed. However it seems there are two neglected issues which are "Cloze and Dictation-type Tests" and "Functional Testing". Considering the fact that syllabus places a high value on these parts of the content, inadequate number of items were observed.

Table 6. The amount of content coverage in Testing (2) exams across teacher training centers

Content	Urmia center		Boroujerd center		Arak center		Shahr-e-Rey center		Tehran centers		Total
Testing Vocabulary	5	13.9 %	9	18.4 %	3	5.9 %	3	6.8 %	6	19.3 %	26
Testing Structure	4	11.1 %	7	14.3 %	8	15.7 %	7	15.9 %	2	6.4 %	28
Testing Pronunciation	5	13.9 %	12	24.5 %	2	3.9 %	5	11.4 %	1	3.2 %	25
Testing Listening Comprehension	6	16.7 %	2	4.1 %	11	21.6 %	6	13.6 %	2	6.4 %	27
Testing Oral Production	2	5.6 %	8	16.3 %	8	15.7 %	3	6.8 %	8	25.8 %	29
Testing Reading Comprehension	3	8.3 %	2	4.1 %	6	11.8 %	7	15.9 %	4	12.9 %	22
Testing Writing Ability	1	2.8 %	4	8.2 %	7	13.7 %	4	9.1 %	3	9.7 %	19
Cloze and Dictation Type Tests	9	25 %	2	4.1 %	2	3.9 %	6	13.6 %	1	3.2 %	20
Functional Testing	1	2.8 %	3	6.1 %	4	7.8 %	3	6.8 %	4	12.9 %	15
Total	36	100 %	49	100 %	51	100 %	44	100 %	31	100	211

Comparing the centers (see Table 6) in terms of content validity of exam questions, Shahid Rajaei of Urmia is in better state than other centers. The questions were evenly distributed among the topics. On the contrary the questions of two centers, namely of Arak and Tehran, were the least valid in terms of content. Quality of questions in Shahid Mofatteh of Shahr-e-Rey is moderate in comparison with other four centers.

To answer the third research question cross-tabulation was conducted that is presented in the table below. From the item classification of 441 questions, five major cognitive process levels along with thirteen sub-levels and three types of knowledge were found. The five major cognitive skills are: Remember, Understand, Apply, Analyze and Evaluate. As for the knowledge type, Metacognitive Knowledge is absent from the list. This neglected area of knowledge in tests designed for assessing prospective teachers' competencies takes on an additional significance while considering its implications for classroom practice. Metacognitive Knowledge is associated closely with some terms such as metacognitive awareness, self-awareness, self-reflection and self-regulation; all can be used interchangeably. It can play an important role in student learning and by implication in the ways students are taught and assessed in the classroom; it seems to be related to the transfer of learning which is the ability to use knowledge gained in one setting or situation in another (Bransford et al., 1999). Teacher educators are to cultivate the awareness of this knowledge among prospective teachers by its inclusion in assessments. One way is to use portfolio assessment that offers students the opportunity to reflect on their work which also leads to gaining self-assessment information. The lack of any question on this knowledge can be attributed to the difficulty of its measurement in formal classroom tests; it is more easily assessed in classroom activities and discussion of various learning strategies (Pintrich, 2002).

Table 7. The crosstabulation of the cognitive sub-categories and knowledge types

Taxonomy	Recognizing	knowledge types			Total	
		Factual Knowledge	Conceptual Knowledge	Procedural Knowledge		
		Count	168	1	0	169
		% of total	38,1%	,2%	,0%	38,3%
	Recalling	Count	167	10	0	177
		% of total	37,9%	2,3%	,0%	40,1%
	Interpreting	Count	2	19	0	21
		% of total	,5%	4,3%	,0%	4,8%
	Exemplifying	Count	0	5	0	5
		% of total	,0%	1,1%	,0%	1,1%
	Classifying	Count	0	5	0	5
		% of total	,0%	1,1%	,0%	1,1%

Inferring	Count	0	12	0	12
	% of total	,0%	2,7%	,0%	2,7%
Comparing	Count	0	7	0	7
	% of total	,0%	1,6%	,0%	1,6%
Explaining	Count	0	8	0	8
	% of total	,0%	1,8%	,0%	1,8%
Executing	Count	0	0	17	17
	% of total	,0%	,0%	3,9%	3,9%
Implementing	Count	0	0	13	13
	% of total	,0%	,0%	2,9%	2,9%
Organizing	Count	0	2	1	3
	% of total	,0%	,5%	,2%	,7%
Checking	Count	2	0	0	2
	% of total	,5%	,0%	,0%	,5%
Critiquing	Count	0	0	2	2
	% of total	,0%	,0%	,5%	,5%
Total	Count	339	69	33	441
	% of total	76,9%	15,6%	7,5%	100,0%

These identified cognitive levels and types of knowledge were then recognized to constitute nine major combinations: (1) Remember Factual Knowledge, (2) Remember Conceptual Knowledge, (3) Understand Factual Knowledge, (4) Understand Conceptual Knowledge, (5) Apply Procedural Knowledge, (6) Analyze Conceptual Knowledge, (7) Analyze Procedural Knowledge, (8) Evaluate Factual Knowledge, (9) Evaluate Procedural Knowledge.

Table 8. The crosstabulation of major cognitive processes and knowledge types

The knowledge dimension		The cognitive process dimension				
		Remember	Understand	Apply	Analyze	Evaluate
Factual Knowledge	Count	335	2			2
	% of total	76 %	0.5 %			0.5 %
Conceptual Knowledge	Count	11	56		2	
	% of total	2.5 %	12.6 %		0.5 %	
Procedural Knowledge	Count			30	1	2
	% of total			6.8 %	0.2 %	0.5 %

As illustrated in tables 7 and 8 among test items classified as Remember (346 in total), 335 of them were identified as testing Factual Knowledge. For items under the category of Understand, 96.5 % of questions (56 items) were labelled as Conceptual Knowledge; remaining 2 items were Factual Knowledge items and none were Procedural Knowledge items. As for items classified as Apply, all of them, comprising about 6.8 % of overall questions, were identified as Procedural Knowledge. Only one sub-category of Analyze, the next cognitive process, was found in the tests which is Organizing. Out of three items, two were measuring Conceptual Knowledge while one tests Procedural Knowledge. For Evaluate both sub-categories, namely Checking and Critiquing, were present and each has two questions. The two items of Checking aim to measure Factual Knowledge whereas the items of Critiquing measured Procedural Knowledge. For the last major cognitive category, Create, no item was found.

4. Discussion & Conclusion

This study was undertaken to investigate what cognitive levels and knowledge types were tested on language assessment literacy exams administered in the past two years at five Teacher Training Centers. The results of item analysis can have a positive washback effect to language assessment literacy classes across teacher preparatory programs. The findings for the first question indicated a narrow range of cognitive processes (i.e. lower-level capacities of Remember and Understand) were mostly being tested in all centers and only a very limited number of questions aim at measuring complex cognitive abilities of Analyze and Evaluate. This is similar to Masters et al. study (2001) in which the majority of the questions were written at the lowest cognitive level of Remember. In the current study after Remember the next cognitive processes in terms of frequency were respectively Understand and Apply corresponding to the order found in Masters et al. study. Also this finding corroborates the results of Squire's study (2001) that found a great proportion of items at Knowledge level of Bloom's Original Taxonomy (corresponding to Remember in Revised Taxonomy) and a few number of items at higher cognitive levels. The result is also compatible with the findings of Rabiavi, et al. (2011). In their study, the exam questions of two teacher training centers were analyzed and they were either labelled as Knowledge or Comprehension on Bloom's Taxonomy; none of the items measured the more complex processes. Current findings is consistent with Lan and Chern's study (2010) that aimed to investigate cognitive process levels and knowledge types measured on the English reading comprehension tests of college entrance examinations in Taiwan. In their study items on Remember Factual Knowledge and Understand Factual Knowledge, which belong to lower cognitive levels, were the majority in the tests under scrutiny and few items were found at higher levels of Apply and Analyze. Analysis of Turkish high-school physics-examination questions by Kocakaya and Gonen (2010) likewise

found only about 27.5 percent of the questions at the higher levels of cognitive domain (analysis, synthesis and evaluation levels). Studies by Hand, Prain and Wallace (2002), Çepni et al. (2003), Karamustafaoglu et al. (2003), Köğçe (2005) also support the view that most traditional examinations are of the lower order cognitive skill type.

In this study, the reason for finding no item on Create category might be due to its productive nature that cannot be grasped through multiple choice or fill-in-the-blanks formats. This view is consistent with Buckles and Siegfried (2006), who found that multiple-choice questions can measure elements of in-depth understanding when being carefully designed, and maintained that Synthesis and Evaluation levels could not be accurately measured since the creativity or originality could not be simply tested via multiple-choice questions.

Concerning the third research question, the following association of cognitive skills with knowledge types surfaced more often than the other combinations in the data set:

- (a) Remember Factual Knowledge
- (b) Understand Conceptual Knowledge
- (c) Apply Procedural Knowledge

As far as the third research question is concerned, the findings are in line with Chen's (2004) in the computer science discipline which revealed this knowledge-and-cognitive association. The result in the present study further confirms Anderson and Krathwohl's (2001) proposition that certain types of knowledge tended to be associated with certain types of cognitive skills.

It is well worth noting that inasmuch as instructors' approach in assessment affects students' learning preferences, the findings of this study have direct implication for classroom practice. As long as the majority of questions only test students' ability to remember or understand the content, the test takers would find memorization and reproduction of knowledge sufficient for getting an acceptable score and would not pay attention to the depth of the content. They would develop a rote and superficial approach to learning without forming a connection between learnings and meaninglessly accumulate information. This situation gets even worse by considering the fact that test-takers will become future teachers and hence become learning models of their own students. Training teachers with this attitude towards assessment adversely affect students since they don't recognize the importance of higher order cognitive process and consequently will not expose their students to these skills. Language assessment literacy course in this sense has got a double significance for future teachers; teacher educators are required to prepare the candidates by familiarizing them with essential knowledge and skills on assessment and getting them to realize the necessity of testing higher level cognitive skills.

Given the above discussion some potential areas for further research emerge. As results of this study demonstrated a gap in the assessment of higher order cognitive skills in teacher preparatory programs, there is a need for developing an alternative evaluation instrument which covers as broad range of cognitive processes as possible and compensates the shortcomings of current assessment procedures. Another issue worthy of investigation is to make sure whether test-takers actually apply expected cognitive skills while answering a question and to check the correspondence between what takes place in reality and the judgment of test-developers. Procedures such as think aloud or interview can be adopted for this purpose. Since "Methodology of Teaching English" is one of the mandatory subjects to be completed towards the English teaching licensure, it is needed to appraise its assessment from cognitive processing perspective.

References

- Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching and assessing: a revision of Bloom's Taxonomy*. New York. Longman Publishing.
- Aviles, C. B. (1999). Understanding and testing for "critical thinking" with Bloom's taxonomy of educational objectives. Washington, DC: ERIC Clearinghouse.
- Bailey, K. M., & Brown, J. D. (1996). Language testing courses: What are they? In A. Cunnig & R. Berwick (Eds.), *Validation in language testing* (pp. 236-256). Clevedon, UK: Multilingual Matters.
- Boyles, P. (2005). Assessment literacy. In M. Rosenbusch (Ed.), *New visions in action: National assessment summit papers* (pp. 18-23). Ames: Iowa State University.
- Bransford, J., Brown, A., & Cocking, R. (1999). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press.
- Brindley, G. (2001). Language assessment and professional development. In C. Elder, A. Brown, E. Grove, K. Hill, N. Iwashita, T. Lumley, T. McNamara & K. O'Loughlin (Eds.), *Experimenting with uncertainty: Essays in honour of Alan Davies* (pp. 126-136). Cambridge: Cambridge University Press.
- Brown, J. D., & Bailey, K. M. (2007). Language testing courses: What are they in 2007? *Language Testing*, 25(3), 349-383.
- Buckles, S., & Siegfried, J. J. (2006). Using multiple-choice questions to evaluate in-depth learning of economics. *Journal of Economic Education*, 37(1), 48-57.
- Çepni, S., Özsevgeç, T., & Gökdere, M. (2003). The comparing questions OSS and entrance high school exam according to cognitive level and the properties formal operational steps. *National Educational Journal*, 157, 1-9.
- Chen, H. J. (2004). An analysis of computer science exam questions using revised Bloom's taxonomy. Unpublished master's thesis, National Taiwan Normal University, Taipei.

- Dochy, F. (2009). The edumetric quality of new modes of assessment: Some issues and prospects. In G. Joughin (Ed.), *Assessment, learning and judgement in Higher Education* (pp. 85-114). New York: Springer.
- Gronlund, N. E. (1985). *Measurement and evaluation in teaching*. New York: Macmillan.
- Hand, B., Prain, V., & Wallace, C. (2002). Influence of writing task on students' answers to recall and higher-level test questions. *Research in Science Education*, 32, 19-34.
- Hüttner, J., Mehlmaur-Larcher, B., & Reichl, S. (2011). *Theory and practice in EFL teacher education: Bridging the gap* (New Perspectives on Language and Education). UK: Short Run Press.
- Inbar-Lourie, O. (2008). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing*, 25(3), 385-402.
- Jin, Y. (2010). The place of language testing and assessment in the professional preparation of foreign language teachers in China. *Language Testing*, 27(4), 555-584.
- Johnson, K., & Johnson, H. (2001). *Encyclopedic dictionary of applied linguistics*. Oxford: Blackwell.
- Karamustafaoglu, S., Sevim, S., Karamustafaoglu, O., & Cepni, S. (2003). Analysis of Turkish high-school chemistry examination questions according to Bloom's taxonomy. *Chemistry Education: Research and Practice*, 4(1), 25-30.
- Kleinsasser, R. C. (2005). Transforming a postgraduate level assessment course: A second language teacher educator's narrative. *Prospect*, 20(3), 77-102.
- Kocakaya, S., & Gonen, S. (2010). Analysis of Turkish high-school physics-examination questions according to Bloom's taxonomy. *Asia-Pacific Forum on Science Learning and Teaching*, 11(1).
- Köğçe, D. (2005). A comparison of the mathematics questions used in the University Entrance Exam (OSS) and high schools in terms of Bloom's Taxonomy, Unpublished master's thesis, Institute of Science, Karadeniz Technical University, Trabzon.
- Kozhageldyeva, L. (2005). *Importance of testing and evaluation in foreign language teaching: A suggested syllable for a testing course in EFL departments in Kazakhstan universities*. Unpublished master's thesis, Gazi University, Ankara.
- Lam, R. (in press). Language assessment training in Hong Kong: Implications for language assessment literacy. *Language Testing*.
- Lan, W. H., & Chern, C. L. (2010). Using Revised Bloom's Taxonomy to analyze reading comprehension questions on the SAET and the DRET. *Contemporary Educational Research Quarterly*, 18 (3), 165-206.
- Masters, J. C., Hulsmeyer, B. S., Pike, M. E., Leichty, K., Miller, M. T., & Verst, A. L. (2001). Assessment of multiple-choice questions in selected test banks accompanying text books used in nursing education. *Journal of Nursing Education*, 40(1), 25-32.
- Mollaenezhad, A. & Zekavati, A. (2007). Comparative study of the curriculum of teacher education systems of England, Japan, France, Malaysia, and Iran. *Educational Innovations Quarterly*, 26 (2), 35-62.
- Newfields, T. (2006). Teacher development and assessment literacy. Authentic Communication: Proceedings of the 5th Annual JALT Pan-SIG Conference. May 13-14, 2006. Shizuoka, Japan: Tokai University College of Marine Science, pp. 48-73.
- Notar, C. E., Zuelke, D. C., Wilson, J. D., & Yunker, B. D. (2004). The table of specifications: Insuring accountability in teacher made tests. *Journal of Instructional Psychology*, 31, 115-129
- Orpwood, G. (2001). The role of assessment in science curriculum reform. *Assessment in Education*. 8(2), pp. 135-151.
- O'Loughlin, K. (2006). Learning about second language assessment: Insights from a postgraduate student on-line subject forum. *University of Sydney Papers in TESOL*, 1, 71-85.
- Popham, W. J. (1987). The merits of measurement-driven instruction. *Phi Delta Kappan*, 68, 679-682.
- Pintrich, P. R. (2002). The role of metacognitive knowledge in learning, teaching, and assessing. *Theory into Practice*, 41(4), 219-225.
- Rabihavi, F., Parsa, A., & Elhampour, H. (2011). Study of compatibility of the exams of teacher training centers of Khuzestan province with psychometry parameters. Paper Presented at Teacher Education Curriculum Conference.
- Shohamy, E. (2005). The power of tests over teachers: The power of teachers over tests. In D. J. Tedick (Ed.), *Second language teacher education* (pp. 101-111). Mahwah, NJ and London: Lawrence Erlbaum.
- Squire, P. J. (2001). Cognitive levels of testing agricultural science in senior secondary schools in Botswana. *Education*, 118(1), 100-110.
- Stiggins, R. J. (1991). Assessment literacy. *Phi Delta Kappan*, 72, 534-539.
- Stiggins, R. J. (1999, Spring). Evaluating classroom assessment training in teacher education programs. *Educational Measurement: Issues and Practice*, 23-27.
- Stiggins, R. J. (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan*, 83(10), 758-765.

Appendix A

The Cognitive Process Dimension

Level 1- C1

Categories & Cognitive Processes	Alternative Names	Definition
Remember		Retrieve knowledge from long-term memory
Recognizing	Identifying	Locating knowledge in long-term memory that is consistent with presented material
Recalling	Retrieving	Retrieving relevant knowledge from long-term memory

Level 2 – C2

Categories & Cognitive Processes	Alternative Names	Definition
Understand		Construct meaning from instructional messages, including oral, written, and graphic communication
Interpreting	Clarifying Paraphrasing Representing Translating	Changing from one form of representation to another
Exemplifying	Illustrating Instantiating	Finding a specific example or illustration of a concept or principle
Classifying	Categorizing Subsuming	Determining that something belongs to a category
Summarizing	Abstracting Generalizing	Abstracting a general theme or major point(s)
Inferring	Concluding Extrapolating Interpolating Predicting	Drawing a logical conclusion from presented information
Comparing	Contrasting Mapping Matching	Detecting correspondences between two ideas, objects, and the like
Explaining	Constructing models	Constructing a cause and effect model of a system

Level 3 – C3

Categories & Cognitive Processes	Alternative Names	Definition
Apply		Applying a procedure to a familiar task
Executing	Carrying out	Applying a procedure to a familiar task
Implementing	Using	Applying a procedure to an unfamiliar task
Analyze		Break material into its constituent parts and determine how the parts relate to one another and to an overall structure or purpose
Differentiating	Discriminating Distinguishing Focusing Selecting	Distinguishing relevant from irrelevant parts or important from unimportant parts of presented material
Organizing	Finding coherence Integrating Outlining Parsing Structuring	Determining how elements fit or function within a structure
Attributing	Deconstructing	Determine a point of view, bias, values, or intent underlying presented material

Evaluate		Make judgments based on criteria and standards	
Checking	Coordinating Detecting Monitoring Testing	Detecting inconsistencies or fallacies within a process or product; determining whether a process or product has internal consistency; detecting the effectiveness of a procedure as it is being implemented	
Critiquing	Judging	Detecting inconsistencies between a product and external criteria; determining whether a product has external consistency; detecting the appropriateness of a procedure for a given problem	

Categories & Cognitive Processes	Alternative Names	Definition
Create	Put elements together to form a coherent or functional whole; reorganize elements into a new pattern or structure	
Generating	Hypothesizing	Coming up with alternative hypotheses based on criteria
Planning	Designing	Devising a procedure for accomplishing some task
Producing	Constructing	Inventing a product

The Knowledge Dimension

Dimension	Definition
Factual Knowledge	The basic elements students must know to be acquainted with a discipline or solve problems in it
Conceptual Knowledge	The interrelationships among the basic elements within a larger structure that enable them to function together
Procedural Knowledge	How to do something, methods of inquiry, and criteria for using skills, algorithms, techniques, and methods
Metacognitive Knowledge	Knowledge of cognition in general as well as awareness and knowledge of one's own cognition

Appendix B

Table of specification for Testing 1 (The Percentage and Number of Questions for Each Section)

Total	Theories	Good test	Construction of items	Statistics	Form s	Function s	Preliminaries s	Content
	14 %	21 %	21 %	14 %	8 %	8 %	14 %	Objective
5	1	1	1	1	-	-	1	Remember 10 %
10	1	2	3	1	1	1	1	Understand 20 %
10	1	2	2	2	1	1	1	Apply 20 %
10	1	2	2	1	2	1	1	Analyze 20 %
10	2	2	2	1	1	1	1	Evaluate 20 %
5	1	1	1	1	-	-	1	Create 10 %
50	7	10	11	7	5	4	6	Total

Instructional Material

Farhady, H., Ja'farpour, A., & Birjandi, P. (2004). *Testing Language Skills: From Theory to Practice*. SAMT Publications: Tehran.

Table of specification for Testing 2 (The Percentage and Number of Questions for Each Section)

Total	Chapter 16	Chapter 15	Chapter 14	Chapter 13	Chapter 12	Chapter 11	Chapter 10	Chapter 9	Chapter 8	Content
	18.5 %	18.5 %	9 %	9 %	9 %	9 %	9 %	9 %	9 %	Objective
5	1	1	-	1	-	-	1	1	-	Remember 10 %
10	2	2	-	1	1	1	1	1	1	Understand 20 %
10	2	2	1	1	-	1	1	1	1	Apply 20 %
10	2	2	1	1	1	1	-	1	1	Analyze 20 %
10	2	2	1	1	1	1	1	1	-	Evaluate 20 %
5	1	-	1	-	1	1	-	-	1	Create 10 %
50	10	9	4	5	4	5	4	5	4	Total

Instructional Material:

Farhady, H., Ja'farpour, A., & Birjandi, P. (2004). *Testing Language Skills: From Theory to Practice*. SAMT Publications: Tehran.