

Exploring the Cognitive Level of Final Exams in Iranian High Schools: Focusing on Bloom's Taxonomy

Saman Ebadi

Assistant Professor, Razi University, Iran

Faranak Shahbazian

MA Student, Razi University, Iran

Abstract

This study investigates the cognitive level of first and second grade of Iranian high-school final exam questions based on Bloom's Taxonomy of Educational Objectives. Cognitive dimension of Bloom's Taxonomy consists of six levels: *knowledge*, *comprehension*, *application*, *analysis*, *synthesis*, and *evaluation*. These levels have a cumulative hierarchy from *knowledge* as the lowest level to *evaluation* as the highest one. A content analysis research was conducted to examine the items and the reliability was assured by establishing inter-coder and intra-coder reliability of the procedure. Content analysis of questions revealed that all the items of first and second grades were at the first three levels of the taxonomy which was mostly the lower order of thinking. The Chi Square test, which was run to investigate the possible specific pattern of frequencies of the items, gave a significant result for both series of questions. It was concluded that there was no specific pattern for the frequencies of the questions. Moreover, there was no difference between first and second grade questions with regard to cognitive levels of Bloom's taxonomy. The results of this study can be useful for raising the knowledge and awareness of test designers and teachers regarding the cognitive level of the items they design. The taxonomy can also be used for planning effective test specifications since it takes both higher and lower cognitive level of items into account.

Keywords: Bloom's taxonomy of educational objectives, higher order thinking, cognitive domain, washback effect, psychometrics

INTRODUCTION

Tests and the goal of the designers

Evaluation and testing have usually been of great importance for students, parents, and teachers due to the important decisions made based on the results of the tests, ranging from passing/failing the school final exams to university entrance exams. Although

playing a great role for accountability purposes, tests are not always merely designed for judgmental decisions. One of the aims for which a test, particularly language test, can be built is encouraging learning. Gipps (1994) states that "educational measurement, by contrast with psychometrics, aims to ... use measurement constructively to identify strengths and weaknesses individuals might have so as to aid their educational progress" (p. 8). Unlike psychometric approach which brings about some limitations on the tests in order to measure only a single attribute at a time, educational measurement attempts to transcend the limitations by using measurement as an instrument for identifying and promoting the abilities of the test takers.

In psychometric approach of testing, the test answers should produce objective and accurate numbers in order for judgmental decisions which are supposed to be made based on the results, so the answers tend to be more close-ended in nature; Interpretive, open-ended questions which need problem-solving skills and higher level of thinking are not usually tapped in this approach. In 1950s, when some educators contended a need for departure from psychometrics, and when the usefulness of psychological measurements in educational settings was doubted, Bloom presented his taxonomy of educational objectives. Bloom's taxonomy has defined six levels for educational objectives whose topmost level is *evaluation* which is believed to be the final objective of education and learning. Gipps (1994) refers to the simultaneity of the emergence of Bloom's taxonomy with the tendency of educators for designing some sort of tests which could focus on educational purposes, not merely on judgmental decisions. She points out that "around the time of the publication of *Bloom's Taxonomy of Educational Objectives* in the late 1950 educators began to articulate a need for assessment which was specifically for educational purposes and could be used in the cycle of planning, instruction, learning, and evaluation. This was termed educational measurement" (pp.7-8).

Bloom's taxonomy of educational objectives

Bloom's taxonomy was created by an educational psychologist, Benjamin Bloom, in order to improve higher order of thinking in education. "The Taxonomy of Educational Objectives is a framework for classifying statements of what we expect or intend students to learn as a result of instruction." Krathwohl (2002, p. 212). Bloom's Taxonomy classifies the objectives that educators set for the learners and it comprises three domains: cognitive, affective, and psychomotor. The first volume of taxonomy which was related to cognitive domain was published in 1956 and followed by the second volume, affective domain, in 1965; however, the most widely used domain of the taxonomy is cognitive domain. The third volume which was planned for psychomotor domain was not published. The six levels of cognitive dimension of this taxonomy consist of *knowledge, comprehension, application, analysis, synthesis, and evaluation*. *Knowledge* is the lowest level of this taxonomy which deals with recalling what the learner has previously learned; this level can be exemplified in definition of facts and principles, or knowing common terms or procedures. *Comprehension* deals with grasping the meaning and interpreting the facts; for example, interpreting the charts and graphs, or translating numbers to

words can be categorized in comprehension level. *Application* is the third level of Bloom's taxonomy which is mainly related to using previously learnt information in new situations to solve the problems; at this level, the learners, for instance, are expected to apply the theories and rules they have learnt before in practical situations or to demonstrate the correct usage of a procedure. *Analysis* is to dismantle a structure in order to investigate the relationships among the separated parts; Distinguishing between facts and inferences is an example of analysis. *Synthesis* level of the Taxonomy is related to producing something new by putting parts together. This level triggers creativity and the ability of formulating new patterns, procedures, and structures. Creating a well-organized short story or poem, or devising a plan for an experiment can be considered as some examples for synthesis level of Bloom's Taxonomy. *Evaluation* refers to the ability of criticism, judgment and making decisions based on definite criteria. The criterion based on which the judgment is made may be given to the learners, or it may be devised by learners themselves. This level of the Taxonomy is believed to be the highest cognitive level which contains all the preceding cognitive levels in addition to the ability of value judgment.

The six levels of Bloom's Taxonomy can be viewed from the perspective of higher-order and lower-order thinking. Higher order thinking includes those kinds of learning that need more cognitive processing but also have more benefits because they prepare students for challenging and real life situations and involves critical thinking. On the contrary, lower order thinking usually includes information that is needed to be recalled rather than being judged, evaluated, or applied. Regarding Bloom's Taxonomy, the three levels of *evaluation, synthesis, and analysis* are considered as higher order thinking, and the *knowledge* and *comprehension* level are considered as lower order thinking. The *application* level can be considered as higher or lower order of thinking depending on its cognitive complexity.

LITERATURE REVIEW

Bloom's taxonomy of educational objectives is one of the most widely referred taxonomies in education and educational planning; it has been used in many fields such as biology, mathematics, engineering, psychology, and so forth. Athanassiou, Mcnett, and Harvey (2003) have used Bloom's taxonomy as a framework in order to study thinking in general education and classroom management. Alavian (2013) investigated the type of thinking that classroom activities developed among EFL learners. Using Bloom's taxonomy for the study, he concluded that there was a difference between the type of thinking available in activities designed by teachers and the type of thinking the teachers aim to achieve in higher level; moreover, it was revealed that the gender and experience of teachers and their level of education was influential in application of thinking activities. Bloom's taxonomy has been used by different researchers for evaluating the EFL textbooks. Razmjoo and Kazempourfard (2012) have evaluated interchange series based on the Bloom's revised taxonomy and concluded that the three low levels in Bloom's revised taxonomy were the most predominant levels which were used in these

books. The higher order thinking levels and also the metacognitive knowledge were absent in interchange series. Gordani (2010) has conducted a content analysis study on Guidance school books based on the Bloom's taxonomy and reached this conclusion that the focus of the books were on the bloom's three lower levels of thinking. Riazi and Mosallanejad (2010) conducted a content- analysis research on the high-school and pre-university books and the result of the study revealed that in these books, again, the focus was only on the lower order thinking levels of the taxonomy; however, the pre-university book was somewhat higher than the other three books regarding to level of thinking.

Nasrollahi (2014) investigated the Iranian EFL students' reaction to teaching critical reading strategies, using cognitive domain of Bloom's taxonomy, and revealed that the critical reading abilities of the students increased and they showed a higher level of thinking. Pourdana and Rajesky (2013) investigated the effect of the difficulty level of EFL texts on the reading performance of EFL learners. They provided 32 undergraduate students majoring in English translation with six different reading texts which were designed based on the six cognitive levels of Bloom's Taxonomy. They observed that as the cognitive level of the texts increases the performance of the learners decreases only except for one level, that is, synthesis which was not statistically supported. The researchers concluded that Bloom's Taxonomy of educational objectives can be a reliable criterion for developing and grading EFL materials because it properly measure learners' performance.

In their article, Luebke and Lorie (2013) presented an account of the use of Bloom's Taxonomy of Educational Objectives by staff of the Law School Admission Council in the 1990 development of redesigned specifications for the Reading Comprehension section of the Law School Admission Test and revealed the usefulness of Bloom's Taxonomy in test specifications and achieving the test goals. Ahmed, Nisa, and Zarif (2013) have elaborated on the development and usage of table of specification in testing and teaching, and the relations between the development of table of specification and cognitive levels of Bloom's Taxonomy. Khorsand (2009) investigated the cognitive level of questions posed by 20 experienced EFL teachers in advanced reading comprehension tests. From the 215 questions posed by these teachers 54.21 percent of the questions were knowledge type, 38.84 percent was comprehension type. Synthesis, application, evaluation, and analysis were respectively 2.33, 1.86, 1.39, and 0.47 percent of the all questions. According to the results of this research only 4.19 percent of the questions were related to the three higher order levels of thinking based on the bloom's taxonomy.

THIS STUDY

This study tries to evaluate the cognitive dimension of nation-wide English final exams of first and second grade of Iranian high schools based on Bloom's taxonomy of educational objectives.

- What are the levels of first and second grade of high-school final-exam questions based on the cognitive dimension of Bloom's taxonomy of educational objectives?
- Which cognitive level of Bloom's taxonomy has been placed more emphasis in these two series of questions?
- What is the difference between the cognitive level of first grade questions and that of the second grade questions based on Bloom's taxonomy of educational objectives?

METHOD

Research design

This research, basically, has a content analysis design. Quantitative information has also applied such as the frequencies and percentages of the items based on the six cognitive levels of Bloom's taxonomy. A chi-square test was run in order to show the possible pattern of the items distribution.

Materials

Materials of this study are the items of English nation-wide final exams of first and second grade of Iranian high school. The final exam of first grade of high school consists of 55 items in 13 parts. The final exam of second grade of high school consists of 49 items in 14 parts. The exams were administered nationwide on June 27th and 29th (2012) at 8.

Data collection and analysis procedures

In order to evaluate the items of the final exams of first and second grade of high school, the researcher has used a codifying scheme. The six levels of Bloom's revised taxonomy have been numbered from one to six with *knowledge* level as number one and *evaluation* level as number six. Then, the researcher evaluated every item of the test based on the six previously defined levels of bloom's taxonomy.

In order to clarify the scheme by which the items have been coded, some sample has been presented. For instance, in the third part of the first- grade high-school exam, the students have been asked to choose the synonyms and antonyms of underlined words; the words have been contextualized within a sentence. All the items in this part have been numbered as 1(i.e., knowledge level) because the students are only asked about their previous knowledge and they are expected just to remember what they have learnt before. Although, the words, in this part, have been contextualized but the whole structure of the question is in such a way that the test takers can easily choose the synonym or antonym of the words without going to the higher level processing of understanding the meaning of the whole sentence.

In twelfth part of this exam the students are asked about the meaning of sentences in the form of a multiple-choice item. In this part the students have to interpret and paraphrase the meaning of the sentences in order to choose the best answer. This part deals with

comprehension and understanding, so all the items of this part have been numbered as 2 (i.e., comprehension level). In another part of the test students are expected to change a declarative sentence to an interrogative one. Because students have to use their previous knowledge to change the sentence in a new form (declarative to interrogative), this item has been coded as number 3 (i.e., application level).

Reliability of the codifying procedure

In order to achieve reliability with regard to codifying of the questions, inter reliability and intra reliability have been taken into consideration. To establish the intra-reliability of the procedure, all the questions of second grade that comprised 47% of the data were codified once more by the researcher after one-month span and the internal consistency of the items was ensured. The degree of consistency between the two codifications was found to be 0.94 based on Cronbach's Alpha. In order to obtain the inter-reliability of the procedure, the same amount of data was codified by two other coders who had M.A degree in TEFL. Before carrying out the codification, they were completely explained about Bloom's Taxonomy and the main goal of the research. The coders were provided with second grade questions and asked to codify the questions based on the six levels of Bloom's taxonomy. To ensure the inter-reliability of coders, Cronbach's Alpha was used and the agreement between the researcher's attempt and that of the other coders was found to be 86.9.

RESULTS

According to the content analysis of the exam questions of first grade of high school, it was revealed that nearly 33 percent of the questions were at knowledge level, 56 percent of the questions were at the comprehension level and, 11 percent of the questions were related to application level.

Table1. The Frequency and Percentage of First Grade Questions

	Frequency	Percentage
Knowledge	18	32.72%
Comprehension	31	56.36%
Application	6	10.90%
Analysis	0	0%
Synthesis	0	0%
Evaluation	0	0%

Table2. The Frequency and Percentage of Second Grade Questions

	Frequency	Percentage
Knowledge	15	30.61%
Comprehension	28	57.14%
Application	6	12.24%
Analysis	0	0%
Synthesis	0	0%
Evaluation	0	0%

Regarding the second grade questions, nearly 12 percent of the questions were related to application of the previous knowledge, 57 percent of the questions were related to understanding the meaning of the text and comprehension, and 30 percent of the questions were at the knowledge level.

In fact, the questions of both grades were mostly focused on the second level of Bloom's taxonomy, that is, comprehension. The other questions were mainly related to the first and third level of the taxonomy which is respectively knowledge and application, all of which are related to the lower order of thinking based on Bloom's taxonomy. The higher levels of thinking have not been taken care of.

In order to investigate the possible specific pattern of items, a chi square test was run for both grades (the significance level is 0.05). The results of the tests have been shown in the table below:

Table3. Chi square test for first and second grade

	First grade	Second grade
Test Statistic	17.055	14.980
df	2	2
Asymp.Sig.	0.000	0.001

Based on the above table, Chi Square test has given us a significant result for both series of questions of first and second grade (Sig=0.000 and 0.001). It can be concluded that the items of different levels of the Taxonomy have been distributed randomly and they do not follow a specific pattern. Moreover, the content analysis of these two series of questions shows that the questions of first grade high school do not significantly differ from those of second-grade high school. That is, the questions of second-grade high school were not at a higher level of thinking as it is logically expected to be.

The questions of these final exams were mostly aimed to elicit the knowledge the students had accumulated before. The learners should try to find the correct answer which has been previously determined. For example, one of the questions of the second-grade high school exam shows a picture in which a lion is in a cage and a boy is watching the lion through the cage bars. The test taker is questioned to answer what would happen if the lion comes out of the cage and the verb *eat* has been put in the parenthesis in order to make the test taker answer the question merely by using this verb *eat* and reach a predetermined expected answer. Furthermore, the students are not allowed to express their own ideas freely by such questions. By changing the question to the following one, the level of thinking can be elevated from lower order to higher order "If you were watching a lion in the cage and the cage door got open, how would you react? Write about your feelings and reactions."

DISCUSSION

The emergence of Bloom's Taxonomy of Educational Objectives in 1956 was around the time that educators tended a departure from psychometrics to a kind of assessment which could be able to support instruction and learning; what was sought was a departure from strict objectivity which was one of the characteristics of psycholinguistic approach of testing. With this regard, the lower levels of Bloom's Taxonomy lend themselves to more objective items; however, by this, lower and higher levels of taxonomy are not intended to be strictly divided into two subjective and objective levels because it is possible for a test designer to design questions in *evaluation* or *synthesis* level which can be scored objectively, but usually higher cognitive questions encourage students to participate in learning processes by critical thinking, personalizing their responses, decision-making ability and self-expression, most of which tend to be evaluated subjectively by raters, not through formulae or strict objective scoring procedures.

In order to elaborate on the consequences of those tests which are at the lower levels of thinking, we take two points into consideration: intended and unintended washback effect discussed by Watanabe and the dilemma of practicality and washback effect suggested by Brown (2004). Watanabe (as cited in Cheng, Watanabe & Curtis, 2004) defined washback effect based on several dimensions including specificity, intensity, length, intentionality, and value. Regarding the value dimension, Watanabe pointed out that "intended washback may normally be associated with positive washback, while unintended washback is related to both negative and positive washback." On the other hand, Brown (2004) convincingly argues that there is a negative correlation between washback effect and practicality of the test.

When the teacher spends more time and effort to offer a reliable evaluation of the students' performance on the test-which is more possible in contextualized, open ended items- positive washback effect (intended or unintended) and intrinsic motivation will be raised among students while at the same time this kind of procedure is likely to reduce the practicality and reliability of the test. Conversely, if the test has been designed to have a higher degree of practicality and reliability, especially in large-scale tests, it is more likely to produce extrinsic motivation and less positive washback effect on the students. If the test designers focus on higher order thinking in designing a test and encourage the test takers to produce more language in test taking, the test may become less practical, but with more positive washback effect.

As indicated in the results, the designers of the first and second grade high-school questions have placed more emphasis on questions which are at the lower level of thinking of Bloom's Taxonomy. One possible reason for such an overemphasis is that the test specification, by which the test designers abide, revolves around lower levels of Bloom's Taxonomy due to relatively more objective nature of these kinds of items and easy scoring procedures especially when the tests are to be administered in large scale.

The test specification may trigger the practicality of the test in advance. If such is the case, there is a need for test specification revision. It is not meant that the lower cognitive questions must be excluded or considered less important than higher level ones. On the contrary, test items should also tap on the lower cognitive ability of the test takers because these levels establish the background knowledge needed for higher order of thinking. The main issue which is to be asserted is the concept of moderation. There should be a moderate weight for both higher and lower cognitive questions and this moderation must not be sacrificed to practicality or easy scoring procedures. When the final exams are merely focused on the lower order of thinking, they may give rise to a kind of language teaching which mainly aims to fit the tests; that is, it may have an unintended negative washback. Although more practical, this kind of test is not likely to develop learning motivation on the part of test takers because they are only expected to choose the correct options or give short, predetermined answers rather than produce open-ended responses involving more higher order thinking and problem solving skills.

The second possible reason for overusing of lower cognitive items in these exams is that the test designers may not have enough knowledge about cognitive dimension of items based on Bloom's Taxonomy. When they construct the items without considering the cognitive levels of items they are more likely to produce items in lower level of cognition due to its concrete and objective nature which makes it easier to be tested. Khorsand (2009) investigated the cognitive level of questions posed by 20 experienced EFL teachers in advanced reading comprehension tests and found that most of the designed questions were at the level of lower order thinking. She asserts that teachers and educators should be trained in order to raise their ability to design items which are at a higher level of cognition. "Teachers (experienced or novice) should be encouraged to attend seminars and teaching method classes which would be designed to enhance ability to reach higher cognitive levels in classroom discourse." (Khorsand, 2009, p.15). Regarding the usefulness of Bloom's Taxonomy, Ahmed, Nisa, and Zarif (2013) point out:

Table of specification is a grid that empowers instructors to develop methodological test instruments. It helps instructor to weigh the different aspects of learning according to blooms taxonomy according to their difficulty level for each module (p. 360).

They believe in the usefulness of devising a table of specification by including Bloom's Taxonomy and refer that Bloom's Taxonomy is applicable even for novice designers due to its comprehensible nature. If we want to raise the ability of thinking among students, we should first focus on raising the higher order thinking ability of teachers. Teachers and test designers should be made aware of the cognitive levels of thinking in general and of Bloom's taxonomy of educational objectives in particular. If they have enough knowledge in this regard, they should apply the knowledge in designing tests and test specifications in order to evaluate students' ability more effectively and encourage the students to think in higher levels; however, by this, we do not mean that the only factor for raising the thinking ability of learners is changing the test items or test specifications.

As Saif (2006) points out that "the test by itself cannot create change in the educational system. There exists an intricate web of different yet related factors that could enhance or interfere with a test's effects being realized as educational change." Yet, the washback effect of tests on teaching and learning should not be underestimated.

CONCLUSION

Based on the results of this study, the focus of English final-exam questions of Iranian high schools is mostly on lower order thinking, and higher order thinking has been, intentionally or inadvertently, neglected in testing to a great extent. Bloom's Taxonomy is one of the best guidelines which can be used by teachers and test designers to bring closer what they teach, what they test, and what they intend to test; it can help teachers and test designers include all the cognitive levels required in effective learning, teaching, and testing without unnecessary overweighing on one level. Teachers can use this taxonomy in their instruction in order to prepare students not only for the exams but also for real life situations and challenging circumstances which invoke critical thinking, decision making ability, creativity, and higher order thinking. Test designers can use this taxonomy for planning better tests and test specifications; although, the inclusion of higher order questions in exams is likely to make the scoring procedures more difficult, but the effort is valuable because by including such questions the evaluation and discrimination of cognitive abilities of test takers would be more reliable and, at the same time, it can produce positive washback effect on EFL instruction.

With regard to Bloom's taxonomy, some suggestions for further research have been made. A different study may be conducted to investigate the cognitive level of English language items of university entrance exams in Iran and a comparison can be made between the level of these series of questions and that of the pre-university English final exam items. This study may help bridge the possible gaps between cognitive levels of thinking the students are practicing in high schools and the cognitive levels they actually are expected to represent in university entrance exams. A separate study can be carried out for evaluating the cognitive levels of thinking existed in reading passages of university entrance exams and that of the pre-university textbooks.

REFERENCES

- Ahmed, A., Nisa, A. & Zarif, T. (2013). Table of specification development and usage: an overview. *Interdisciplinary Journal of Contemporary Research Business*, 4(12), 354-361.
- Alavian, S.J. (2013). The impact of Iranian EFL classroom activities on the development of desired thinking process as suggested by Bloom's Taxonomy. *Frontiers of Language and Teaching*, 4, 14-33.
- Athanassiou, N., Mcnett, J.M., & Harvey, C. (2003). Critical thinking in the management classroom: Bloom's taxonomy as a learning tool. *Journal of Management Education*, 27(5), 533-555.

- Brown, H. D. (2004). *Language assessment: Principles and classroom practices*. White Plains, NY: Pearson Education.
- Gipps, C. V. (1994). *Beyond testing: Towards a theory of educational assessment*. London: The Falmer Press.
- Gordani, Y. (2010). An analysis of English textbooks used at Iranian guidance school in terms of Bloom's Taxonomy. *The Journal of Asia TEFL*, 7(2), 249-278.
- Khorsand, N. (2009). Cognitive levels of questions used by Iranian EFL teachers in advanced reading comprehension tests. Retrieved May 2, 2014 from <http://eric.ed.gov/?id=ED507869>.
- Krathwohl, D. (2002). A revision of Bloom's taxonomy: An overview. *Theory Into Practice*, 41(4), 212-218. Retrieved May 2, 2014 from http://www.unco.edu/cetl/sir/stating_outcome/documents/Krathwohl.pdf
- Luebke, S., & Lorie, J. (2013). Use of Bloom's taxonomy in developing reading comprehension specifications. Retrieved May 2, 2014 from <http://0-eric.ed.gov.opac.msmc.edu/?q=Use+of+Bloom%e2%80%99s+taxonomy+in+devel+oping+reading+comprehension+specifications.+&id=EJ1013203>.
- Nasrollahi, M.A. (2014). Iranian EFL learners' reactions to the process of teaching critical reading strategies: using Stringer's action research model. Retrieved May 2, 2014 from <http://ssrn.com/abstract=2491051>
- Pourdana, N., & Rajesky, J.S. (2013). Estimating the difficulty level of EFL texts: Applying Bloom's Taxonomy of Educational Objectives. *International Journal of Applied Linguistics & English Literature*, 2(6), 202-211.
- Razmjoo, S.A., kazempourfard, E. (2012). On the representation of Bloom's revised taxonomy in interchange course books. *The Journal of Teaching Language Skills*, 4(1), 171-204.
- Riazi, A. M., & Mosallanejad, N. (2010). Evaluation of learning objectives in Iranian high-school and pre-university English textbooks using Bloom's Taxonomy. *TESL EJ*, 13(4), 1-16.
- Saif, S. (2006). Aiming for positive washback: a case study of international teaching assistants. *Language Testing*, 23 (1), 31-41.
- Watanabe, Y. (2004). Methodology in washback studies. In L. Cheng, Y. Watanabe & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods*. Mahwah, N.J: Lawrence Erlbaum.