

Original Paper

Levels of Behavior: Do we Really Test Higher Skills at Higher Levels?

Netra Prasad Sharma¹ (Note 1) & Kamal Kumar Poudel¹ (Note 2)

¹ Tribhuvan University, Department of English Education, Mahendra Ratna Campus, Tahachal, Kathmandu, Nepal

Received: December 25, 2020

Accepted: January 6, 2020

Online Published: January 9, 2021

doi:10.22158/wjer.v8n1p94

URL: <http://dx.doi.org/10.22158/wjer.v8n1p94>

Abstract

Although opinions often vary regarding what makes higher education inherently different from lower levels of education, it is generally agreed upon the fact that, among other parameters, the former is virtually characterized by higher levels of cognitive behavior on the part of students. Employing the theoretical framework of Bloom and associates (1956) and the subsequent revisions of the framework, this paper examines eight question papers meant for testing the achievements of students in two courses of study taught at the Graduate level in English education. Upon a careful examination of the questions from the perspective of Bloom et al.'s taxonomies, the data reveal that the focus of assessment either lacks clarity or basically lies on measuring the lower-order skills, a fact that goes counter to the very general assumption of higher education. Drawing from the findings, it is recommended that the concerned authority should review the extant assessment practices in line with the advocacy of the nature of higher education.

Keywords

levels of behavior, higher education, question papers, higher-order skills, lower-order skills

1. Introduction

Ever since Bloom et al.'s (1956) publication of the taxonomies of educational objectives, the discourse of what level(s) of cognitive behavior one focuses on during teaching and assessing one's students has prevailed, in fact, worldwide. In the case of the Nepali educational institutions, this kind of discourse is getting substantial space relatively recently both in school as well as university level educational assessment systems. The focus of such discourse, at least in most cases, previously remained limited mainly to planning lessons and teaching rather than assessing students. Some recent concrete attempts of National Examinations Board towards standardizing its tests have also geared up the discussion on

the related issues like specificity of test items in terms of the levels of behavior they intend to measure, the dissenting remarks of some academics regarding the quality of the tests, and so on. Furthermore, add-on remarks though, some publications also point out the issue of cognitive levels. The general argument of some academics that the university level tests are also limited to testing mainly the lower-order cognitive skills is what this study intends to pursue.

1.1 Context

Higher education, also known as tertiary education, is generally conceived to be the level at which higher-order skills are developed through intensive instruction and practices so as to prepare students for employability and lifelong learning. Such skills are supposed to equip those students with the ability to cope with the novel situations they might encounter even in the post-formal-education phase of their lives. Thus, tertiary education that is pursued after the elementary and secondary education, carries a paramount significance for those who follow it and the society at large. It is, therefore, often portrayed as something that may not be within the reach of everyone and, hence, rather an optional matter than compulsory. It is taken to be a sort of privilege not accessible to everyone because of several reasons such as cognitive, socio-economic, and the like.

In the Nepalese context, for the reasons to be explored, the number of higher education entrants getting through secondary education seems to be remarkable. Nonetheless, the expectations of the Nepalese society from the university Graduates turn out to result in increasing unemployability and a massive flow to the foreign lands every year even for the work that only slightly, if any, harmonizes with university degrees. This leads to a grave question against the quality of higher education – the question for university academics being “What are we teaching and testing in higher education?”. This question is further propped by the frequent reactions from Nepalese academics that we are imparting the store of theoretical knowledge that can be manipulated merely in habitual situations and that has almost no significance for solving the problems university Graduates might encounter in their real-life worlds. It is normally believed that the teaching-and-learning process in different streams of university education, not all in fact, is limited largely to knowledge and comprehension, and that it rarely goes beyond that. This, as we believe, is true about the case of English language education in the Nepalese context as well, and, therefore, our interest in this study.

1.2 Objective

The objective of the study was to explore the levels of skills tested at the Graduate level of English language education in Tribhuvan University. In so doing, the study attempted to sensitize the concerned stakeholder groups to the need of constructing test items and tasks in the spirit of higher education, the spirit of generally focusing on the higher-order skills rather than the lower-order ones for which only the retrieval of information and comprehension suffice.

1.3 Research Questions

This study attempted to answer the following research questions in order to attain the objective:

- What are the levels of skills tested at the Graduate level of English language education?

- How well do the question papers of English language education carry the spirit of higher education?

2. Review of the Related Literature and Theoretical Framework

As an attempt to pulling together the previous works – both theoretical and empirical – to a short glance, this section reviews the literature relevant to the study in an integrated manner. Yet, they have been conceived as comprising two thematic headings, namely, focus of higher education and levels of behavior to be tested, and presented accordingly in the text that follows.

2.1 Focus of Higher Education

Hussey and Smith (2010) assert that “learning outcomes must differ according to the level of teaching and learning concerned. ... what is appropriate at primary school is quite different from what is required at university” (p. 58). In the present world dominated by “knowledge economies”, higher education learning is conceived to comprise the inherent features such as multidimensional and dynamic strategic processing, self-regulation, learner adaptation, metacognition, critical thinking, judgment and decision-making skills, problem solving skills, and so on. For instance, on sampling some articles, Alexander (2017) arrives at some constructs of higher education – “metacognition, self-regulation, self-regulated learning, student approaches to learning, strategic processing, deep-level and surface-level strategies, motivated learning strategies, engagement, motivation, perceived control, and need for competence” (p. 356). Additionally, she also mentions the “habits of mind that support professional success and lifelong learning” (p. 345) as the constructs of higher education. “Higher Education institutions spur human resource skills-enrichment, which in turn builds the capacity to compete in a globalized world in which the Knowledge Economy reigns supreme” (Al-Hawaj, 2008, p. IX).

Similarly, Mathema, and Bista (2006) highlight the value of higher-order skills thus:

The use of tests emphasizing lower-order abilities promotes learning strategies that are superficial or short-term (memorizing, rehearsing, and rote learning). A system of education that only aims to develop such abilities on the part of learners has little to offer towards the realization of the individual and social goals of education – economic growth, nation-building, social transformation, and development of creative and independent citizens. Demand for higher level skills will grow further as more and more youths are seeking employment outside the country. (p. 414)

Nicholls (2002) points out the common failings of the assessment in higher education, some of them being “too heavy a reliance on subjective judgment”, “hasty assessment preparation”, “use of short, inefficient assessment tools”, “testing trivia”, “careless wording of assessment questions”, failure to analyze the quality of the test” (p. 109), and so on. He opines that the assessment procedures need to reflect the learning processes, the failure in which results in the fact that “academics include assessment strategies that tend to examine what students have memorized rather than whether they can apply, analyze and critically reflect on what they have learned” (p. 108). This suggests why the focus on

lower-order cognitive skills results in assessment procedures.

It is to be noted that the reporting of ideas and opinions presented in the foregoing text is meant not to suggest, in any way, that the assessments in higher education should aim at measuring the higher-order skills only. It is also remarkable at this point that the levels of skills to be taught and assessed are not at all delimited to and determined by factors such as age of students, their formal academic levels as elementary, secondary or tertiary, and the like. Likewise, it is also true that the courses of academic disciplines do make a difference in the definition of such skills, i.e., a skill defined as a lower-order skill in a specific context of a course might be conceived as a higher-order one in another. In the same way, a skill treated as a higher-order one for a group of students might be a lower-order one for a different group. Similarly, even the elementary and secondary level learners need to practice higher-order skills to the extent within their control. This clearly suggests that there are levels of skills within a skill. For instance, the level of “application” for elementary students is virtually different from the level of “application” for the students at the tertiary level. Therefore, the point in our case is that of focus and the level of depth required of learners in learning and assessment. This implies that the emphasis on the assessment of just the lower-order skills, such as recall and comprehension, at the tertiary level cannot at all be justified, and, therefore, it does not assist in attaining the goals of higher education, the goals aspiring to prepare learners for solving real-world problems through their higher-order skills and also to equip them with the tactics for life-long learning. In the light of the increasing complexity of Bloom et al.’s (1956) cognitive levels and the generally expected behavior of the students in higher education, it is almost definite that the focus of higher education should be on higher-order skills rather than on the lower-order ones. The concentration is, in general, required on deep-level approaches, metacognition, dynamics of high level cognitive processing such as critical thinking, multidimensionality, self-regulation, and so forth.

2.2 Levels of Behavior

Although the idea of taxonomizing the educational objectives for teaching, learning and assessing was germinated “at an informal meeting of college examiners attending the 1948 American Psychological Association Convention in Boston” (Bloom, Engelhart, Frust, Hill, & Krathwohl, 1956, p. 4), it was materialized so as to concretely inform the world only in 1956 with the very publication by Bloom et al. (1956). The contents of the publication are popularly known as “Bloom’s taxonomy” even though it is not only Bloom who contributed to the exposition of the framework. Then, it is also important to note that originally the idea emerged in order to facilitate the communication among examiners – the fact indicating the origin of the idea of taxonomy in the motivation for better assessment. However, in our context, the framework seems to be discussed more in teaching and learning rather than testing and assessment. Bloom et al.’s (1956) theoretical framework for the classification of educational objectives consists of three domains, viz. cognitive, affective and psychomotor. Nonetheless, this paper focuses on the first of the domains only.

The original framework forwarded the six levels for the cognitive domain – the levels being knowledge,

comprehension, application, analysis, synthesis, and evaluation. However, there have been a number of revisions in the taxonomy during and after the 1990s. One of the major revisions is Anderson, Krathwohl, Airasian, Cruikshank, Mayer, Pintrich, Rath, and Wittrock (2001) in which the original labels have been renamed and the position of the last two levels has been reorganized. The revision employs the terms “remember”, “understand”, “apply”, “analyze”, “evaluate” and “create”. In using the terms in verbal forms, the authors hold that, as opposed to the nominal forms carrying the meaning of the knowledge dimension as a product, as used in the original version; verbal forms carry in them the nuance of cognitive processes, which, in effect, emits an impression of an ongoing entity. There also exists the practice of using the progressive forms of the verbs representing the cognitive levels (Churches, 1997). The progressive forms suggest more explicitly the ongoing nature of the cognitive levels. Additionally, the levels of behavior have been defined in the order of increasing cognitive complexity from knowledge/remember to evaluation/create. The following section illuminates the levels and their basic features.

2.2.1 Remembering

This level comprises the retrieval of the information already stored in the minds of learners/examinees requiring them of recognizing and recalling ideas, materials, events or any phenomena. It lays emphasis on rote learning. The verbs such as count, list, recall, state, define, name, recite, tabulate, draw, point, recognize, identify, read, repeat, memorize, reproduce, etc. can be used in wording test items.

2.2.2 Understanding

This level of cognitive behavior basically calls for the comprehension of ideas. Organization of ideas, finding relationships between/among ideas, manipulation of knowledge, etc. are the focus of this level. Anderson et al. (2001, p. 31) list interpreting, exemplifying, classifying, summarizing, inferring, comparing and explaining as the skills that can be placed at this level. The verbs such as associate, describe, interpret, differentiate, estimate, translate, classify, discuss, compare, translate, interpret, explain, identify, locate, recognize, report, select, paraphrase and so on can be used while phrasing test items and tasks.

2.2.3 Applying

This level of behavior requires, on the part of learners/examinees, the skill in using the previously acquired/learned knowledge and their comprehension of material in concrete situations. As the term suggests, one has to be able to apply one's knowledge and comprehension in order to solve problems in novel situations – those that are different from the ones already familiar and in which learners have previously practiced the materials in question. Apply, demonstrate, practice, use, calculate, relate, examine, operate, sketch, solve, dramatize, illustrate, design, etc. are some of the verbs that can be used while wording test items.

Usually, this level is divided into two sublevels when the issue of categorizing the levels of behaviour into either higher-order levels or lower-order ones is raised. Bakker (2018) discusses the terms “production” and “reproduction” with reference to HO-LO test items. The two terms, in effect,

represent the two sublevels of “application”. The first sublevel of this intermediate level between HO and LO is the level in which learners are required to apply the already learnt knowledge in the situations that they are well familiar with. The similar situations as the ones in which learners have had enough practice of materials require less cognitive complexity and, therefore, less cognitive pressure on the part of the learner. This level is, therefore, categorized as the lower-order cognitive level. The second sublevel is, however, categorized as the higher-order cognitive level in that it anticipates much more powerful and complex cognitive performance in order to solve problems by applying the extant knowledge and comprehension of learners/examinees. Situations are novel/unique and the materials to be manipulated are the ones not already practised. Learners are required to adopt some creative procedures rather than the well-trained ones and also that they may have to create some material on their own as well.

2.2.4 Analysing

This level of behaviour requires higher-order skills. This requires learners/examinees to “break material into constituent parts and determine how parts relate to one another and to an overall structure or purpose” (Anderson et al., 2001, p. 31). The authors subsume the three skills, viz. differentiating, organizing and attributing under this level. The verbs such as analyze, break down, compare, contrast, deconstruct, differentiate, separate, relate, etc. can be used in wording test items.

2.2.5 Evaluating

This level was placed by Bloom et al. (1956) at the topmost position of cognitive complexity. However, the revised versions have swapped the positions of “evaluation” and “synthesis” and also that the label “synthesis” has been renamed as “creating”, the level that appears at the top in place of “evaluation” in the original version. This level entails “judgment based on criteria and standards” (Anderson et al., 2001, p. 31). This includes “checking” and “critiquing”. The test items at this level can be worded using verbs like appraise, critique, defend, evaluate, interpret, justify, argue, support, etc.

2.2.6 Creating

This level appears at the topmost position in the revised versions and requires the skill to “put elements together to form a coherent or functional whole; reorganize elements into a new pattern or structure” (Anderson et al., 2001, p. 31). Thus, this level, in essence, carries the contents of “synthesis” in the original version of the taxonomies. Bloom et al. (1956) state that:

Generally this would involve a recombination of parts of previous experience with new material, reconstructed into a new and more or less well-integrated whole. This is the category in the cognitive domain which most clearly provides for creative behavior on the part of the learner. (p. 162)

Test items can be phrased employing the action words such as combine, compile, compose, reconstruct, modify, and the like.

The levels of cognitive behavior described in the foregoing paragraphs in this section have been conceived to fall into two major categories, viz. higher-order skills and lower-order ones. Remembering,

understanding and the first sublevel of application have been subsumed under the lower-order skills, whereas the second level of application, analyzing, evaluating and creating have been placed under the higher-order category. It was done for the sake of preciseness of data presentation.

When it comes to empirical literature, a study from Turkey carried out by Karamustafaoğlu, Sevim, Karamustafaoğlu and Cepni (2003) reports that 96% of the items in “chemistry” question papers intended to test the lower-order cognitive behavior. However, as the study reveals, more than 50% of the test items in a university entrance examination intended to test higher-order cognitive behavior. In a similar vein is Kocakaya and Gönen’s (2010) study from Turkey that analyzed the physics question items employed in schools and reveals that 72.5% of the test items fell into the lower-order category. Upon comparing the results with the levels of questions used in a university entrance exam, however, 50.9% items fell into the higher-order category. The authors state that this discrepancy between high school level assessment and that of the university creates problems. In the same way, Razmjoo and Madani (2013) from Iran analyzed the test items of a university entrance exam in terms of Bloom’s revised taxonomy and revealed that the items intended to measure lower-order skills more than higher-order ones. The study also revealed that there was the “complete absence” of “creating” level items. The authors assert that the exam “cannot make learners critical thinkers” (p. 83). Yet, another study comes from Nepal. The study undertaken by Martin Chautari (2018) reports that 76% of the test items in the Compulsory English test employed in the School Leaving Certificate level fell into the category of lower-order skills. Mathema and Bista (2006) also report somehow similar results of the analysis of the six core subjects. “Analysis of test papers used in SLC in six core subjects suggests that the test items, for the most part, are designed to test the acquisition of lower-order abilities at the cost of higher-order abilities” (p. 414), the researchers report. They conceive the fact to be the “most disturbing” one. Newstead and Hoskins (2003) briefly report the research findings arrived at in Entwistle and Entwistle (1991) and Newstead and Findlay (1997) that, at the beginning, students work intrinsically with deep learning approaches but, later, they switch over to surface and rote learning due to some extrinsic factors – a rather similar case, as we believe, in the context of higher education in Nepal. The authors suggest that “one way of changing this might be if the assessment system were to be one which encouraged conceptual understanding as opposed to rote learning” (p. 71).

The empirical evidence presented in the foregoing paragraph clearly suggests the rather disapproving results. The focus of the test items was laid fundamentally on the lower-order skills encouraging superficial rote learning and mere comprehension. In this scenario of test item focus, to our knowledge, no studies have used the framework as the present study did to extensively analyze the test items employed on the tertiary level English language education tests in Tribhuvan University. This study, therefore, locates itself in this void and aspires to address the issue under investigation.

2.3 Theoretical Framework for the Study

This study was carried out within the framework originally proposed by Bloom et al.’s (1956) and the subsequent revisions of the framework for the classification of learning and assessment objectives,

which has been delineated in the immediately preceding section.

3. Method

In order to answer the research questions set for this study, the qualitative approach to data gathering and analysis was adopted. In particular, following the document analysis strategy, eight question papers administered by the Faculty of Education from 2013 to 2016 to assess the achievements of the students at the Graduate level were analyzed in terms of the framework proposed by Bloom and his associates in 1956 and the revisions thereafter. The question papers consisted of four, each of two courses of study purposively selected as the sample based on our familiarity with the courses, relevance and the representation of the courses of different nature. For instance, the question papers of Course I and Course II were chosen in that the question items of the former are, in general, meant for testing the higher-order skills as “critical thinking” and the course also represents the “content courses”, whereas, the latter represents the courses pertaining to “pedagogy”, usually considered to be the very heart of the Faculty of Education.

Each of the test items of the eight question papers was first judged independently and carefully by the principal author and assigned to a category based on the levels of behavior the item in question supposedly intended to measure. Then, as a check for the trustworthiness of the judgment, the subjective questions of each of the two courses were shared with two teachers, including the corresponding author, teaching the respective courses. Thus, based on the framework we delineated and also shared with the teacher supposed to carry out the rating, the teacher teaching Course I had to categorize 60 questions of the course whereas the corresponding author had to categorize only 56 as the number of the questions in the test papers of those two courses was different. The categorization of the test items has been displayed using appropriate tables and the conclusion of the study drawn in line with the results and discussion of the data.

4. Results and Discussion

Judgment was made in terms of the two categories of cognitive behavior, viz. higher-order behavior and lower-order behavior. The broad categories encompass in them the six levels of behavior as outlined by Bloom and his associates. For the sake of preciseness of data presentation, knowledge, comprehension and the first level of application were subsumed under lower-order behavior whereas the second level of application, analysis, synthesis and evaluation under higher-order behavior.

Table 1. Ratings of Questions by Principal Author

Course I													Remarks
Q	2013			2014			2015			2016			
	HO	LO	HO/LO	HO	LO	HO/LO	HO	LO	HO/LO	HO	LO	HO/LO	
1.		√			√		?			√			1. In Course I, objective items based on reading texts mostly at the scanning level and, therefore, intending to measure lower-order skills. In Course II, objective items mostly requiring the recall of information imparted by the teacher or reported in prescribed textbooks and/or other non-authentic materials available in the market. 2. In subjective section of Course I, lots of exact liftings in the question papers of 2013 and 2014. 3. HO= higher-order questions; LO= lower-order questions; Q= question
2.	√				√			√			√		
3.		√			?		?					√	
4.		√		√			?			?			
5.		√			√			?		√			
6.		√		√			√			√			
7.			√		√			√		?			
8.	?				√		?				?		
9.		√			√		?			?			
10.		√		√			√			?			
11.	√				√		√				?		
12.		√			?			√		?			
13.	?			√			?			?			
14.		√		?				?		√			
15.		√		?									
16.		√		?									
Course II													
Q	2013			2014			2015			2016			
	HO	LO	HO/LO	HO	LO	HO/LO	HO	LO	HO/LO	HO	LO	HO/LO	
1.	√				√		√				?		
2.		?		√				√			√		
3.		√			√		√			√			
4.		?			?		?				?		
5.		?			√			?				√	
6.		?			√			?			√		
7.			√		√			?			?		
8.		√			√			?			√		
9.		√		?				?			√		
10.		√			√			?			√		
11.	√				√			√			√		
12.	?			?					√		√		
13.		?		?				√			√		
14.	?			√			?			√			

Table 1 shows the way the principal author generated the data for analysis. There appeared different types of test items based on his judgment. First, the items represented by “ticks” indicate that the items clearly fall either under the category of “higher order” or “lower order”. For instance, the question “*Do you agree with the statement ‘most students’ formal education has little connection to real life?’ Explain your answer with reference to Nepal*”, even though based on the reading text by John Holt included in the prescribed textbook prepared by Gardner (2009), is clearly a higher-order question in that it requires examinees to relate the ideas they have read to the Nepali context, which demands, beside recall and comprehension, the skills for analysis, evaluation and creation.

The ticks under the “HO/LO” category suggest that the test items in question contain two or more components and that one or more component questions are clearly either higher-order or lower-order questions. For example, the question “*What do you mean by inductive approach to teaching grammar? Prepare three different inductive exercises to teach the Past Simple tense in English*” comprises two components, the answer to the first can be supplied based only on recall and comprehension also as the description of “inductive approach” is available in several references and the non-authentic market materials like guidebooks and guess papers. Nonetheless, the second component is obviously, at least, at the second level of “application” or, say, even the “creating” level.

Similarly, the question marks represent several items which, as the first author judged, are difficult to clearly categorize either as higher-order items or lower-order ones for the reason that looking at the items as such they seem to fall obviously into one of the categories as the question marks under HO or LO indicate but, when the author considered the test items in terms of the prescribed textbooks, they are either exact liftings or their wording only slightly modified – requiring more or less the same contents as required by the textual questions or the texts, which the target students must have virtually practiced well enough previously. Thus, such questions require mainly the retrieval of the information the students have stored in advance – necessitating almost nothing to analyse, synthesize or evaluate. Therefore, the question marks mean that some of the test items are essentially the ones in which the students have been trained well earlier but, due to some wording changed, they give the false impression of being related to higher-order items. This makes the clear categorization of the items either into one category or the other difficult – resulting in the possibility of the items changing their class and even allowing space for the contradictions between two or more researchers. For instance, the question “*An American friend of yours is visiting your country. What should he/she know in order to avoid intercultural misunderstanding?*” is, in effect, a textual question (Gardner, 2009, p. 13) reading “Explain what a visitor to your country should know in order to avoid intercultural misunderstandings”. Here, firstly, the question, when compared with the textual one, is obviously a lifting from the prescribed textbook with the slight modification of wording only, and, therefore, students must have adequately practiced answering it previously. Secondly, the reading text entitled “American Values and Assumptions” presents the American cultural artefacts, which are enough for examinees of the Graduate level to compare the Nepali cultural artifacts with. Thus, even though the question seems, at a

quick glance, to belong to the HO category, it is, in fact, a lower-order question that can be answered completely based on recall and comprehension.

A good number of the question marks in Table 1 reveal rather an interesting-but-odd fact that the question setters undertaking the sensitive task like question setting – the task that impacts on the destinies of hundreds of thousands of examinees, are either devoid of any clear idea as to exactly what level of behavior they intended to assess through the test items in question or they ignored the devastating consequences of fuzzy test items lacking any clear direction and objective. The reasons behind this await further explorations.

From the category-wise comparative perspective, a total of 116 subjective questions included in the selected question papers were rated independently first by the principal author and, then, by the corresponding author as well as the teacher teaching Course I. Table 2 portrays the number of questions under each category.

Table 2. Number of Questions Under Each Category

	HO	LO	HO?	LO?	HO/LO
PAR	21 (18.10%)	45 (38.79%)	24 (20.68%)	21 (18.10%)	5 (4.31%)
TR	41 (35.34%)	48 (41.37%)	14 (12.06%)	10 (8.62%)	3 (2.58%)

Legends: HO = clearly higher-order, LO = clearly lower-order, HO?= not clear and could be marked LO as well, LO? = not clear and could be marked HO as well, HO/LO= component questions clearly either of higher- or lower-order, PAR= Principal author rating, TR= teacher rating.

As the data show, nearly 40% of the questions represented by “HO?” and “LO?” lack any clear goal as to the level of cognitive behavior. A good number of questions rated by two teachers also indicate the similar fact. The percentage of clearly HO and LO, as indicated in the first two columns is about 57% in case of the first author’s rating whereas considerably more in case of teacher rating (i.e. almost 77%) and the number of questions in the “mixed category” of HO and LO, as shown in the last column, is remarkably low.

Upon a close observation, three facts come to the surface (Table 2). Firstly, despite the variation in PAR and TR in terms of individual questions, there is a clear pattern in the data on the whole – the pattern portraying the tendency that, comparatively, the concentration of test questions is on the lower-order items than on the higher-order ones. Secondly, the data along the columns “HO?” and “LO?”, both in PAR and TR, indicate the lack of clarity on the part of test constructors as to what level of behavior they intend to measure. Thirdly, the figures in the last column, small though, suggest that test constructors intend to measure two or more levels of questions through the same question resulting in a

mess in the holistic scoring adopted in marking answer sheets. The revelations obviously go counter to the expectations of higher education as discussed under the section “focus of higher education” above. The answer to the question “Do we really test higher skills at higher levels?” is, as the data prop, is more of “no” rather than “yes”. The results of this study, therefore, align with the ones as reported in Karamustafaoğlu, Sevim, Karamustafaoğlu, and Cepni (2003), Kocakaya and Gönen (2010) and Kocakaya and Gönen (2010) – the studies reporting the inclusion of higher-order items in university level tests to be only around 50% or less.

Despite the limitations of the study such as relatively a smaller sample size resulting from various reasons and some disparity in the PAR and TR that could be further narrowed down by intensive discussion among raters, the findings closely tallying with those of other similar studies render the trustworthiness of this study. Large-scale studies involving the multiple courses of study, different levels of university education and even more raters might be useful for further substantiating the trustworthiness of the findings.

5. Conclusion

Stemming from the researchers’ encounters with the comments made in the academic circle that the test items, even at the higher levels in Tribhuvan University, mostly belong to the lower-order category of cognitive levels, this paper attempts to scrutinize the comments of the academics by analyzing the subjective questions of eight test papers employed by the University to measure the achievements of the Graduate level students. The facts reveal that, firstly, the comments of the academics are stuffed to a large extent; and, secondly, the analysis of the questions disclose an interesting finding that the question setters are either not sure about what they are intending to measure through their questions or they intend to use the same question to assess the abilities and skills of students at different cognitive levels – the fact evident in the question marks in Tables 1 and 2 and causing a sort of fuzziness in scoring examinee responses, and hence, further threatening the reliability of scoring the subjective responses. What all this suggests is the need for the review of the whole process of setting questions in the University, including the training of question setters. Furthermore, this paper conclusively suggests that in setting questions, the question setters need to consider the spirit of higher education as well.

References

- Alexander, P. A. (2017). Issues of constructs, contexts, and continuity: Commentary on learning in higher education. *Educ Psychol Rev*, 29, 345-351. <https://doi.org/10.1007/s10648-017-9409-3>
- Al-Hawaj, A. Y. (2008). Welcome message. In A. Y. Abdulla, E. Wajeel, & E. H. Twizell (Eds.), *Higher education in the twenty-first century: Issues and challenges .Proceedings of the international conference, Ahlia University, Kingdom of Bahrain, 3-4 june 2007* (pp. IX-X). London: Taylor & Francis Group.
- Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R.,

- Raths, J. & Wittrock, M. C. (Eds.) (2001). *A taxonomy for learning, teaching and assessing: A revision of Bloom's taxonomy of educational objectives* (abridged edition). Longman: New York.
- Bakker, S. (2018). *Slides shared in Item Writing Training organized by NEB in Kathmandu on 18-22 July 2018*.
- Bloom, B. S., Engelhart, M. D., Frust, E. J., Hill, W.H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives, handbook 1, cognitive domain*. USA: David McKay Company, Inc.
- Churches, A. (1997). *Bloom's digital taxonomy*. Retrieved from <http://burtonslifelearning.pbworks.com/f/BloomDigitalTaxonomy2001.pdf>
- Hussey, T., & Smith, P. (2010). *The trouble with higher education: A critical examination of our universities*. New York: Routledge.
- Karamustafaoğlu, S., Sevim, S., Karamustafaoğlu, O., & Çepni, S. (2003). Analysis of Turkish high-school chemistry-examination questions according to Bloom's taxonomy. *Chemistry Education: Research and Practice*, 4(1), 25-30. <https://doi.org/10.1039/B2RP90034C>
- Kocakaya, S., & Gönen, 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), 1-14.
- Martin, C. (2018). *Measurement of cognitive abilities in the SLC examinations in Nepal*. Research Brief No. 21. Retrieved from <http://www.martinchautari.org.np>
- Mathema, K. B., & Bista, M. B. (2006). *Study on student performance in SLC*. Kathmandu: Ministry of Education and Sports & Education Sector Advisory Team.
- Newstead, S. E., & Hoskins, S. (2003). Encouraging student motivation. In H. Fry, S. Ketteridge, & S. Marshall (Eds.), *A handbook for teaching & learning in higher education* (pp. 62-74). London: Kogan Page.
- Peter, S. G. (2009). *New directions: Reading, writing, and critical thinking*. New Delhi: Cambridge University Press.
- Razmjoo, S. A., & Madani, H. (2013). A content analysis of the English section of university entrance exams based on Bloom's revised taxonomy. *International Journal of Language Learning and Applied Linguistics World*, 4(3), 83-103.

Notes

Note 1. Dr. Sharma, principal author, Lecturer (English Education), Email: tu.netrasharma@gmail.com

Note 2. Mr. Poudel, corresponding author, Associate Professor (English Education): Email: kkpoudel2023@gmail.com