Journal of Psychological Research
http://ojs.bilpublishing.com/index.php/jpr

ARTICLE

Beyond Bloom's: Students' Perception of Bloom's Taxonomy and its Convolution with Cognitive Load

Andrea J. Phillips Jace C. Briggs Jamie Lee Jensen*

Department of Biology, 4102 LSB, Brigham Young University, Provo, UT 84602

ARTICLE INFO

Article history
Received: 2 December 2018
Accepted: 26 February 2019
Published: 30 April 2019

Keywords:
Bloom's taxonomy
Cognitive load theory
Student perception
HOCS
LOCS

ABSTRACT

Bloom's taxonomy is widely used in educational research to categorize the cognitive skills required to answer exam questions. For this study, we analyzed how students categorize exam questions (high-level question or low-level question,) gathered data as to their rationale for categorization, and compared their categorizations to those of experts. We found that students consistently rank high-level questions incorrectly. We analyzed student reasons for their categorizations, and found that for many of the incorrectly categorized questions the students referred to reasons related to Cognitive Load Theory. This shows that cognitive load prevents students from accurately assessing the cognitive level of an exam question. Thus, extra cognitive load in exam questions may prevent those questions from accurately measuring the skills and knowledge of the student. This points to the need for instructors to eliminate cognitive load from their exams.

1. Introduction

A guiding framework by which to create assessment items is Bloom's Taxonomy of the Cognitive Domain [1]. Bloom's Taxonomy was originally designed to allow assessment practitioners to speak a common language when referring to assessment items and/or the learning goals of a course or curriculum[2]. Originally published in 1956 as the Taxonomy of Educational Objectives: The Classification of Educational Goals, Handbook I: Cognitive Domain [3], it was intended to be a cumulatively hierarchical system by which to rate the cognitive processes used to solve any given assessment item.

In 2001, Anderson and Krathwohl [2] revised the taxonomy to reflect a more active approach to the cognitive skills, replacing the nouns in the original taxonomy with verbs (see Figure 1). In addition, they created a two-dimensional table incorporating both knowledge and the cognitive process domains to facilitate the categorization of assessment items (see Table 1). Typically, educational professionals refer simply to the cognitive process dimension when categorizing items. In fact, it has become common practice within most major textbooks in the sciences to label assessment items at the ends of chapters or in test banks and online tutorials using this system.

Bloom's levels are often associated with difficulty in relation to the cognitive tasks required to successfully complete the item. It is commonly accepted that the first two levels of Bloom's (Remember and Understand) require very little understanding and cognitive activity. These are commonly referred to as lower-order cognitive skills (LOCS); whereas, the application level and above require deeper conceptual understanding and higher-order cognitive skills (HOCS) to solve [4,5]. Difficulty, in this

**Corresponding Author:*
Jamie Lee Jensen,
Department of Biology, 4102 LSB, Brigham Young University, Provo, UT 84602
Email: Jamie.Jensen@byu.edu

DOI: https://doi.org/10.30564/jpr.v1i1.421

respect, however, is often conflated with performance [6,7]. Some researchers have shown that students perform better on lower-level items and worse on higher-level items [8,9]. However, student performance on an item frequently does not reflect the actual cognitive difficulty assigned by Bloom's Taxonomy [10].

Some have suggested that perhaps other criteria can contribute to the overall difficulty or cognitive demand of assessment items, from the processes involved to the cognitive level of the concepts being assessed. Nurrenbern and Robinson [11] suggest that familiarity with processes may play a role in item difficulty. They proposed that questions be divided into three levels: those that require simply recalling information (Recall), those that require the student to apply a common or familiar calculation or formula (Algorithmic) and those that require students to blend several pieces of information and transfer them to a new situation (Higher-order).

In regard to the content knowledge being assessed, Lawson and others [12] have suggested that there are varying levels of concepts that increase in difficulty hierarchically, the acquisition of which is highly dependent on a students' cognitive developmental level. The first level is descriptive and consists of concepts whose meanings come directly from experience, e.g., environmental factors, food chains, and populations. The second level is hypothetical and consists of concepts whose meaning may be theoretical but could be observed directly if the time frame was extended, e.g., species, limiting factors, and evolution. The third level is theoretical and consists of concepts whose meaning cannot be directly observed but must be implied from indirect observations, e.g., osmosis, genes, and photosynthesis. Lawson et al. found that as the level of concept increased, the knowledge of this concept among students decreased. Thus, the difficulty of an item may be influenced by the level of the conceptual materials that must be recalled to perform at any of the levels of Bloom's Taxonomy.

The purpose of this research was to determine how students categorize the Bloom's level of exam questions, as well as to gather qualitative data as to their methodology for categorization. We also wanted to compare student ratings of question difficulty to their actual performance on the questions. From this data, we hoped to answer the following questions:

1) Do students and experts rate Bloom's items the same?

2) Does questions difficulty correspond to student or instructor ratings?

3) Do students offer any additional explanation for their ratings?

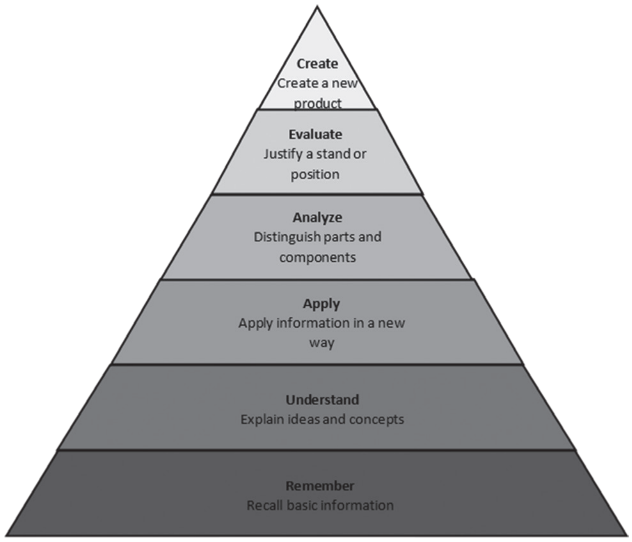

Figure 1. The Revised Bloom's Taxonomy of Educational Objectives.

Table 1. Two-dimensional assessment table

Knowledge Dimension	Cognitive Processes Dimension					
	Remember	Understand	Apply	Analyze	Evaluate	Create
Factual						
Conceptual						
Procedural						
Meta-cognitive						

2. Methods

2.1 Ethics Statement

We were granted permission to use human subjects in this study our institution's Institutional Review Board for Human Subjects. Written consent was provided by all students participating in this study.

2.2 Participants

Participants for this study, ranging from freshman to senior undergraduates, were recruited from an introductory, non-science majors' biology course that is part of the General Education Requirements of the University and is taken by all non-science majors. This study was conducted over three semesters, for a total of 30 groups of students, consisting of 2 – 3 students each; responses were reported by consensus of the group, hence 30 collective responses were recorded. Count of groups was recorded, but to maintain anonymity, the exact student count and student roll was discarded after credit was given.

 DOI: https://doi.org/10.30564/jpr.v1i1.421

3. Research Design

The research was designed to answer three research questions:

1. Do students and experts rate Bloom's questions the same?

2. Does Question Difficulty Correspond to Student or Instructor Ratings?

3. Do Students Offer Any Additional Explanation for Their Ratings?

To answer these questions, students were asked to analyze the final exam of the course they had just finished, prior to receiving any feedback, and offer their interpretations of the difficulty of the questions and the thinking skills required to answer them.

3.1 Procedure

Students participated in the entirety of the course that was taught with student-centered pedagogy[13], and met 3 times a week for 50 minutes each class period. Each unit of the class began with exploratory activities, followed by vocabulary word introduction and concept application activities. Following the final exam but before the students had received feedback (i.e., they had not received their scores), students were recruited to participate in this study. Students voluntarily participated, with the incentive of a small amount of extra credit and free food.

The students were divided into small groups of 2-3, and were given instructions on how to complete the given task, including a brief description of the difference between declarative (content-based) and procedural (skill-based) knowledge[14], along with a few examples of content questions and the skills needed to answer those questions. The students were not given a description of Bloom's taxonomy or the difference between HOCS/LOCS to avoid influencing student answers. Students were instructed to examine each exam question, and report on the following criteria for each item:

a. What content was required and what was the difficulty of that content (easy vs. hard)

b. A description of why they classified the content as easy or hard

c. What thinking skills were required (what your brain had to do) to answer the question

d. Whether those thinking skills were easy or hard

e. A description of why they classified the skill as easy or hard

An example of the student rating sheet is shown in Table 2.

Table 2. Student rating sheet provided to the student volunteers.

		Content (Concepts and Terms)			Thinking Skills (What your brain had to do)			
Item #	List	Easy	Hard	Why?	List	Easy	Hard	Why?
1								
2								
...								

After completing their rating task, students were thanked for their time and given their extra credit. No further contact was made with students.

3.2 Measures

3.2.1 Final Exam

The final exam for the course was cumulative, with no time limit, and no outside resources permitted. It was administered during the last week of the semester in the university testing center. While the concepts covered in the exam were covered during the semester in both class activities and exams, the items on the final exam were unique. The exam contained 41 multiple-choice questions over the content. Twenty-one of these questions required higher-order cognitive skills (HOCS,) and 20 of these questions required lower-order cognitive skills (LOCS) as ranked by experts. Expert ratings (HOCS/LOCS) were decided by a set of three researchers trained in the categorization of Bloom's levels [15]. Raters discussed each question until they came to 100% agreement on its ranking.

3.2.2 Bloom's Ratings

Four independent researchers analyzed student rating cards to recategorized thinking skills listed by students into either LOCS for easy thinking skills (Remembering-Understanding) or HOCS for hard thinking skills (Applying-Creating). Categorizations were discussed until agreement was reached. These independent researchers consisted of the lab principal investigator (PI) and three undergraduate researchers trained in Bloom's Taxonomy and in educational research designs over the course of six months. This group of independent researchers also established the 'expert' rating using Bloom's Taxonomy as a guide.

3.2.3 Item Difficulty

The difficulty of each exam question was defined by the proportion of students who answered the question correctly (i.e., the higher the number, the less difficult the ques-

DOI: https://doi.org/10.30564/jpr.v1i1.421

tion). Students' perception of difficulty was determined by the number of groups that rated the item as easy divided by the total number of groups, i.e., the higher the number, the easier they perceived the question to be.

3.2.4 Additional Explanations

We used content analysis to explore the different explanations given by students (explaining the reasons why they rated the skills required by a question as easy or hard). Explanations were reviewed by the three independent undergraduate researchers mentioned previously. Explanations that related directly to the Bloom's taxonomy level of the question were excluded, as we were interested in the reasons given that were not related to Bloom's taxonomy level. The remaining explanations were grouped according to common themes. The independent researchers, along with the lab PI, then came together to compare common themes.

4. Statistical Analyses

4.1 Do Students and Experts Rate Bloom's Questions the Same?

To answer our first question, student ratings (LOCS or HOCS) were compared to expert ratings to determine agreement. A Chi-squared analysis was used to determine how many groups needed to agree to not be due to random chance. It was found that 21 out of 30 groups needed to agree with experts to be statistically considered full agreement on the cognitive level of a question. From there, bins were created to represent level of agreement from fully agreed to fully disagreed with experts on the question categorizations. These bins are shown in Table 3, along with the number of groups required to fall into each category.

Table 3. Number of groups required to be considered at a certain level of agreement with expert ratings.

Level of Agreement with Expert	Number of Groups
Agree	21 or more
Somewhat Agree	17-20
Ambiguous	14-16
Somewhat Disagree	10-13
Disagree	9 or less

4.2 Does Question Difficulty Correspond to Student or Instructor Ratings?

To answer our second question, the difficulty of each exam question was determined. The relationship between the actual question difficulty and the student perception of question difficulty was determined by running a Pearson's product-moment correlation and a paired-samples t-test. Then, a logistic regression was run to determine whether the question difficulty (as determined by actual student performance) predicted expert ratings of Bloom's level (HOCS/LOCS).

4.3 Do Students Offer Any Additional Explanation for Their Ratings?

To answer our third question, themes recorded by more than one reviewer were kept, and their frequency recorded. Themes related to cognitive load theory were most frequently mentioned, so a logistic regression was performed to determine whether the mention of themes related to cognitive load were predictive of the expert categorization of the question's Bloom's level. Then, an ANOVA was conducted to determine whether mention of cognitive load was related to the student perception of content difficulty (easy vs. hard). A second ANOVA was conducted to determine whether mention of cognitive load was related to student perception of skill level (HOCS/LOCS). Finally, a third ANOVA was conducted to determine whether mention of cognitive load was related to the actual question difficulty as determined by student performance.

5. Results

5.1 Do Students and Experts Rate Bloom's Questions the Same?

To answer our first question, we compared student ratings of Bloom's level (HOCS or LOCS) with expert ratings (Figure 2). We found that student and expert ratings were in higher agreement on LOCS questions than HOCS questions. We found that of the 20 expert-rated LOCS items, students were in agreement on 19 of them. However, on the 21expert-rated HOCS questions, students only agreed on 9 of them, rating many of them as LOCS items.

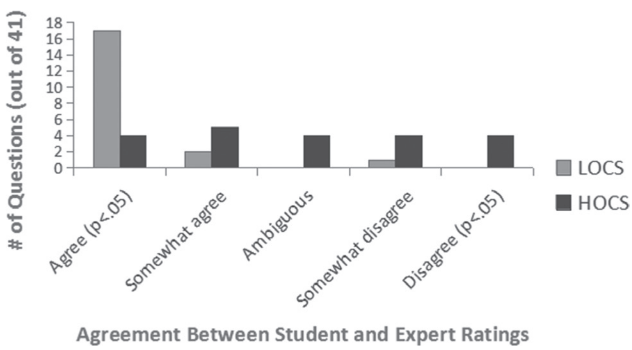

Figure 2. Level of agreement between ratings (higher-order HOCS or lower-order LOCS) given by students and experts, and the number of questions falling into each category out of 41 total questions.

5.2 Does Question Difficulty Correspond to Student or Instructor Ratings?

To answer our second question, we ran a Pearson's product-moment correlation to determine the relationship between student perception of content difficulty and their actual performance on the exam. Question difficulty in this case is defined by the proportion of students who answered the question correctly. We found a linear relationship with a moderate positive correlation between content difficulty and student performance, $r(39)= 0.393$, $p<0.05$, with perceived question difficulty explaining 15.4% of the variation in student performance (Figure 3). A paired-samples t-test shows that students rated questions as significantly less difficult (M=0.66, SD=0.25; i.e., the percentage of groups that rated the item as easy) than their actual performance (M=0.52, SD=0.24), $t(40)=3.23$, $p=.003$.

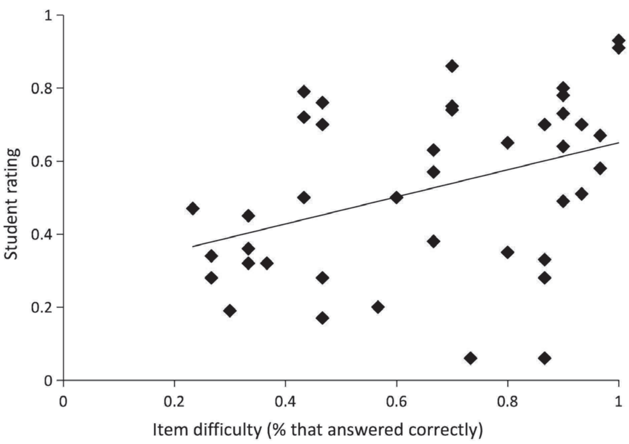

Figure 3. Relationship between student rating of question difficulty, and the actual question difficulty based on how many students answered the question correctly.

A logistic regression was run to determine whether the question difficulty (determined by student performance) predicted expert ratings (HOCS/LOCS). Or in simpler terms, we wanted to determine whether there is a correlation between the "difficulty" of a question and its Blooms categorization. The logistic regression model was statistically significant χ^2 (1) = 4.23, p = 0.04. However, the model explained only 13.1% (Nagelkerke R2) of the variance in HOCS/LOCS classification and correctly classified only 65.9% of questions.

5.3 Do Students Offer Any Additional Explanations for Their Ratings?

To answer our third question, we reviewed the students' explanations behind their categorizations of difficulty in skills needed to answer the question to look for any common themes. The themes mentioned are summarized in Table 4. The responses given by students could be grouped into 3 categories: difficulty related to cognitive load [16], difficulty due to multiple correct answers [6], and ease due to familiarity [4,17].

Within the category of difficulty due to cognitive load, there were 4 common subthemes that related to cognitive load theory, which we will illustrate with student quotes. The first was that the question required extensive reading or took a long time to complete. One group commented that an item "takes time to understand what is going on and at this point, [my] brain is kind of shut-down." Another commented that an item was a "longer, wordy question." A second theme was that the question required sorting through relevant and irrelevant information. One group commented that an item "requires an ability to separate important and non-important information." Another group commented that an item "took too long and required separation of important info and fluff." And another said that the item "included information that wasn't necessary which was confusing." A third theme was that the question required multiple steps to solve. One group commented that the item required students to "understand multiple diagrams [and] go back and forth between questions and diagrams". Another group said that there were "a lot of steps involved, not just memorization." And another group said, "There's a lot of steps we had to take." A fourth theme was that the question required the recall of multiple concepts. One group made the comment that to solve an item, you "had to combine a lot of concepts to know the answer." Another group commented that, in order to solve the problem, you "had to know properties of a lot of topics making you have to think about chemistry." And another group said that an item required them "to apply a lot of principles."

Students who found the item difficult due to multiple correct answers made comments such as, "[the] directions made it hard to distinguish which answer could be right," or "[we] couldn't quite understand the difference between the choices," or "options were all really close to the answer; had to choose 'most correct' answer." Those who found the item easy due to familiarity said things like, "It was covered in class and all needed info was easily studied", or "we used oxygen a lot in class as an example" (referring to a problem involving oxygen's electronegativity), or "[it was easy] because they were defined in class [or] learned in high school."

6. Discussion

6.1 Summary of Findings

We set out to answer three research questions by asking students to analyze and offer feedback on the final exam,

DOI: https://doi.org/10.30564/jpr.v1i1.421

Table 4. Reasons behind student ratings of the cognitive skill level of exam questions, and the proportion of HOCS/ LOCS questions where each reason was mentioned.

Student Categorization	Frequency of HOCS questions (Total = 20)	Frequency of LOCS questions (Total = 20)	Number of groups that mentioned it (Total = 30)
Difficulty due to cognitive load	18	2	20
Extensive reading/time	16	0	26
Relevant/irrelevant information	10	1	13
Multiple steps to solve	10	1	13
Recall of multiple concepts	14	1	16
Difficulty due to multiple correct answers	16	5	26
Ease due to familiarity	13	18	20

specifically commenting on the difficulty of items and the cognitive skills required to solve them.

6.2 Do Students and Experts Rate Bloom's Questions the Same?

Our results show that students consistently rate LOCS the same as experts. As LOCS questions tend to require only recall from students, it follows that students should be able to accurately rate these questions as LOCS, even if they are unable to choose the correct answer. However, students did not show the same consistency in their ranking of HOCS questions, nor do their ratings of HOCS align with expert ratings. Student inability to correctly identify the cognitive skills required in HOCS questions may be due to the higher cognitive level of the questions themselves. Additionally, students' incorrect perception of the cognitive level of a question is arguable due to cognitive load. Longer passages, multiple steps, or the integration of multiple concepts may lead students to believe that a question requires higher cognitive skills than it does. On the other hand, students may incorrectly rank HOCS questions as LOCS due to the familiarity of a question. Formulaic questions may be perceived as lower-order by students because they have been frequently exposed to those question types during class.

6.3 Does Question Difficulty Correspond to Student or Instructor Ratings?

Before discussing this question, it is prudent to reassert the distinction between Bloom's level and the difficulty of a question. While Bloom's rating (HOCS or LOCS) corresponds to the level of the cognitive skills required to correctly answer the question, question difficulty refers to the difficulty of the content being tested. Therefore, the difficulty of the question refers to the number of students who answer the question correctly. While this distinction is important, it is worthy of mention that HOCS questions tend to be more difficult than LOCS questions by virtue of the higher cognitive skills required to answer them correctly.

Our data shows that students consistently rated questions as being easier than they were. Previous research supports that students are often overconfident about their abilities[18]. Upon analyzing the difficulty of test questions, we found that question difficulty does not correspond to the students' HOCS/LOCS categorization of that question. Considering the distinction between Bloom's level and question difficulty discussed above, this is not surprising.

Contrary to student ratings, we found that expert ratings of Bloom's level did correspond to question difficulty, though it did not account for much variation in student performance. Only 13% of the variation in student performance on test questions was accounted for by the Bloom's level of test questions. The reasons for this finding bring us to our final research question.

6.4 Do Students Offer Any Additional Explanation for Their Ratings?

When asked why they ranked questions as HOCS or LOCS, students gave several reasons for their ratings, falling into the following categories: High-level for reasons related to Cognitive Load Theory (i.e. there was a lot of reading involved, both relevant and irrelevant information was presented, the question required multiple steps, or the question required integration of multiple concepts)[16]; High level because there were multiple potentially correct answers presented (multiple answers could be argued for) [6]; or low-level because the question was algorithmic—in other words, familiar questions seemed easier to answer [4,17].

6.5 Theoretical Implications

There have been multiple taxonomies or classifications used in analyzing the cognitive level of assessment items. Bloom's Taxonomy [3] is one of the most salient classifications in the literature. Bloom's Taxonomy was originally designed to allow assessment practitioners to speak a common language when referring to assessment items and/or learning goals of a course. The aim of the taxonomy was to classify items based on the cognitive processes required by the item. It was originally intended to be hierarchical in terms of complexity [19]. However, some have suggested that only the first three levels are hierarchical, while analyze, evaluate, and create are not [4]. It was not, however, intended to be a measure of difficulty. In fact, some categories may overlap with higher categories having easier questions on occasion than lower categories. This is well supported by empirical evidence [19].

Zoller [5] simplified the taxonomy, grouping memorization and recall items into the broad category of lower-order cognitive skills (LOCS) and application, analysis, and evaluation of concepts as higher-order cognitive skills requiring a deeper conceptual understanding. It has been suggested by some that HOCS tend to be more difficult items [8,9] while others have shown little association between level of question and student performance [21]. This suggests that other factors, other than the level or complexity of the cognitive skill, may be at play.

Lemons and Lemons [6] investigated this issue from a faculty perspective and found several additional factors that faculty perceived to influence the 'level' of a question that went beyond Bloom's Taxonomy. Those items included difficulty, the time required to solve the problem, student experience with the problem, and how many correct or semi-correct answers were presented.

Our study suggests that, from a student perspective, there are indeed other factors at play in influencing the complexity of an assessment item. Our students perceived three factors that influence item difficulty: cognitive load, multiple correct answers, and familiarity. Cognitive load refers to the mental procedures happening during the learning process [22] but it is not often applied to the mental procedures occurring during test taking. We would suggest that the addition of a cognitive load component to a taxonomy of assessment items would allow a greater depth of classification. Students in our study offered several characteristics of a problem that could contribute to overall cognitive load: extensive reading, the need to sort through relevant and irrelevant information, the requirement of multiple steps to solve a problem, and the need to recall several concepts at once.

Other studies have shown similar findings. Several have shown that the time required to complete a task is often associated with difficulty [23], or that context or framing of an item can play a large role in its difficulty [24-26], or that items requiring students to reason across levels of organization cause more difficulty [27,28] With regard to familiarity, cognitive load theory would suggest that if a student has a schema for the particular problem, it becomes much less difficulty[23]. Mesic and Muratovic [27] suggest that the automaticity of knowledge structures can play a role in difficulty, those items with higher automaticity being more familiar and therefore easier.

6.6 Practical Implications

Our results should draw attention to the fact that there are many other factors, beyond Bloom's Taxonomy, that should be considered when writing assessment items. Instructors can often introduce additional cognitive complexity unintentionally by increasing the cognitive load of a question with complex scenarios or the additional of superfluous information. Certainly, it is not a bad thing to teach students to wade through information to find what is salient, to be able to quickly identify necessary data, to make sense of a potentially convoluted scenario, or to hold multiple concepts in their working memory at the same time. These are all skills that are necessary in real life situations where complex problems are never straight-forward. However, to achieve alignment between assessments and learning outcomes, it is necessary for instructors to recognize the additional skills that are necessarily built in to their assessment items related to cognitive load and to explicitly (or implicitly) teach these skills as part of their class instruction.

To make this alignment process easier, we would suggest the need for a new taxonomy that combines Bloom's Taxonomy skills with factors of cognitive load, levels of familiarity, and complexity of the problem context. With this new taxonomy, instructors can better define expected learning outcomes and design assessment items to align with these.

6.7 Limitations

A few limitations should be taken into consideration when interpreting these data. First, responses were gathered in groups, rather than from individual students. We hope that this means that responses were a consensus of the group; however, it is possible that one or two individuals in a group dominated the opinions of others. Thus, the actual proportion of responses may be less representative of the whole group. Further study should be done on a larger proportion of students. In addition, the institution where this data was collected is a highly selective private institution and as such represents a group of students that may not be comparable to open-enrollment public institutions. Items that were perceived as easy at this institution may have been perceived as more difficult at other institutions. Interestingly, our students still consistently ranked items

  DOI: https://doi.org/10.30564/jpr.v1i1.421

easier than their actual performance showing an over-estimation of their abilities, something common to most students. It would be fruitful to run a similar study at more institutions representing a broader range of student bodies. Lastly, student responses to prompts were often brief and not in full sentence form. Depth could have been added to these responses by being able to interview students and get further clarification on their answers. A video transcript of student discussions in the future may lead to deeper and more meaningful breadth to form a better taxonomy.

8. Conclusion

The reasons given for students' rankings help us to account for the remaining variation in student performance on exam questions. As instructors, we wish to assess the content knowledge of our students and their ability to apply complex cognitive skills to the content, not their ability to perform well on exams when cognitive load is a major factor. Eliminating excess cognitive load from exams allows instructors to accurately assess conceptual understanding. Alternatively, perhaps it is our goal to help students manage the cognitive load associated with finding reliable information or solving complex problems in today's world. In this case, it behooves instructors to be aware of and more explicit about these additional learning goals. Future research may offer valuable insight into how the presence of cognitive load impacts the accuracy of student ratings of Bloom's level. Additionally, a modified classification framework that takes cognitive load into account when categorizing exam questions may be prudent to offer greater accuracy to both our analysis of exam questions, and the ability of exams to accurately measure student conceptual knowledge.

References

[1] Bloom, B. S. (1984). Taxonomy of educational objectives. [M] Boston: Allyn and Bacon.

[2] Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. [J] Theory into Practice, 41, 4: 212. doi:10.1207/s15430421tip4104_2

[3] Bloom, B., Engelhart, M. D., Furst, E. J., Hill, W. H., Krathwohl, D.R. (1956). Taxonomy of Educational Objectives: The Classification of Educational Goals, Handbook I: Cognitive Domain [M]. New York: David McKay.

[4] Crowe, A., Dirks, C., Wenderoth, M. P. (2008). Biology in Bloom: Implementing Bloom's Taxonomy to Enhance Student Learning in Biology. [J] Cbe-Life Sciences Education, 7, 4: 368-381. doi:10.1187/cbe.08-05-0024

[5] Zoller, U. (1993). Are lecture ad learning compatible? Maybe for LOCS: Unlikely for HOCS [J]. Journal of Chemical Education, 70, 3: 195-197. doi: 10.1021/ed070p195

[6] Lemons, P. P., Lemons, J. D. (2013). Questions for Assessing Higher-Order Cognitive Skills: It's Not Just Bloom's. [J] Cbe-Life Sciences Education, 12, 1: 47-58. doi:10.1187/cbe.12-03-0024

[7] Wyse, A. E., Viger, S. G. (2011). How item writers understand depth of knowledge. [J] Educational Assessment, 16, 4: 185–206. doi: 10.1080/10627197.2011.634286

[8] Knecht, K. T. (2001). Assessing cognitive skills of pharmacy students in a biomedical sciences module using a classification of multiple-choice item categories according to bloom's taxonomy. [J] American Journal of Pharmaceutical Education, 65, 4: 324-334.

[9] Freeman, S., Parks, J. W. (2010). How accurate is peer grading? [J] CBE Life Sciences Education, 9, 4: 482–488. doi: 10.1187/cbe.10-03-0017

[10] Momsen, J. L., Long, T. M., Wyse, S. A., Ebert-May, D. (2010). Just the Facts? Introductory Undergraduate Biology Courses Focus on Low-Level Cognitive Skills. [J] Cbe-Life Sciences Education, 9, 4: 435-440. doi:10.1187/cbe.10-01-0001

[11] Nurrenbern, S. C., Robinson, W. R. (1998). Conceptual questions and challenge problems. [J] Journal of Chemical Education, 75, 11: 1502. doi: 10.1021/ed075p1502

[12] Lawson, A. E., Alkhoury, S., Benford, R., Clark, B. R., Falconer, K. A. (2000). What kinds of scientific concepts exist? Concept construction and intellectual development in college biology. [J] Journal of Research in Science Teaching, 37, 9: 996–1018. https://doi.org/10.1002/1098-2736(200011)37:9<996::AID-TEA8>3.0.CO; 2-J

[13] Armbruster, P., Patel, M., Johnson, E., Weiss, M. (2009). Active Learning and Student-centered Pedagogy Improve Student Attitudes and Performance in Introductory Biology. [J] Cbe-Life Sciences Education, 8, 3: 203-213. doi:10.1187/cbe.09-03-0025

[14] deJong, T., FergusonHessler, M. G. M. (1996). Types and qualities of knowledge. [J] Educational Psychologist, 31, 2: 105-113. doi:10.1207/s15326985ep3102_2

[15] Anderson, L. W. (2002) This issue – Revising Bloom's taxonomy. [J] Theory Into Practice, 41, 4: 210-211. doi:10.1207/s15430421tip4104_1

[16] Sweller, J. (1988). Cognitive load during problem-solving – effects on learning. [J] Cog-

nitive Science, 12, 2: 257-285. doi:10.1207/s15516709cog1202_4

[17] Nurnberger, J. I., Blehar, M. C., Kaufmann, C. A., Yorkcooler, C., Simpson, S. G., Harkavyfriedman, J., . . . Ritz, A. L. (1994). Diagnostic interview for genetic-studies – rationale, unique features, and training. [J] Archives of General Psychiatry, 51, 11: 849-859. PMID: 7944874

[18] Bell, P., Volckmann, D. (2011). Knowledge Surveys in General Chemistry: Confidence, Overconfidence, and Performance. [J] Journal of Chemical Education, 88, 11: 1469-1476. doi:10.1021/ed100328c

[19] Anderson, L. W., Krathwohl, D. R., and Bloom, B. S. (2001). A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives, New York: Longman.

[20] Anderson, L. W., Krathwohl, D. R., and Bloom, B. S. (2001). A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives, New York: Longman.

[21] Momsen, J., Offerdahl, E., Kryjevskaia, M., Montplaisir, L., Anderson, E., Grosz, N. 2013. Using assessments to investigate and compare the nature of learning in undergraduate science courses. CBE-Life Sciences Education 12, 239-249.

[22] Kotovsky, K., Hayes, J.R., Simon, H.A. (1985). Why are some problems hard? Evidence from Tower of Hanoi. Cognitive Psychology (17), 248– 294.

[23] Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. Cognition and Instruction, 12, 185–233.

[24] Hammer, D., Elby, A., Scherr, R.E., Redish, E.F. (2005). Resources, framing, and transfer. In: Transfer of Learning from a Modern Multidisciplinary Perspective, ed. JP Mestre, Greenwich, CT: Information Age, 89–120.

[25] Nehm, R.H., Ha, M. (2011). Item feature effects in evolution assessment. Journal of Research in Science Teaching (48), 237–256.

[26] Nehm, R.H., Schonfeld, I.S. (2008). Measuring knowledge of natural selection: a comparison of the CINS, an open response instrument, and an oral interview. Journal of Research in Science Teaching(45), 1131–1160.

[27] Bloome, D., Puro, P., Theodorou, E. (1989). Procedural display and class- room lessons. Curriculum Inquiry (19). 265–291.

[28] Jime´nez-Aleixandre, M.P., Bugallo Rodr´ıguez, A., Duschl, R.A. (2000). "Doing the lesson" or "doing science": argument in high school genetics. Science Education (84), 757–792.

[29] Mesic, V., Muratovic, H. (2011). Identifying predictors of physics item difficulty: a linear regression approach. Physical Review Physics Education Research (7)010110.

DOI: https://doi.org/10.30564/jpr.v1i1.421