

# The Application of Bloom's Taxonomy to Higher Education Examination Questions in Physics

Jessica Gates & Samantha L Pugh\*

School of Physics and Astronomy, University of Leeds, UK

\*Corresponding author: [s.l.pugh@leeds.ac.uk](mailto:s.l.pugh@leeds.ac.uk)

**Keywords:** *Physics; Exams; Bloom's Taxonomy*

---

### Abstract

Examination papers were analysed using a methodology based on Bloom's Taxonomy to identify the cognitive skills required to complete questions and compare these to the cognition necessary for graduate skills. This research found that examinations assess mainly mid to low-level cognition such as recall and apply, while competencies required by employers tend to need higher-level cognition such as synthesis and creation, which are not as commonly tested through examinations. This paper proposes that careful design of examination questions using different measurable verbs could be more effective at encouraging development of higher-level metacognitive skills in formal examinations.

### Introduction

The core curriculum for undergraduate physics is well established (The Quality Assurance Agency, 2017), and is verified through an accreditation process by professional bodies, such as the Institute of Physics (IOP) in the UK (Institute of Physics, 2010).

Assessment in Physics Higher Education is predominantly conducted by way of formal examinations, with many traditional physics modules relying on exam results to determine the majority of a grade (Fry et al, 2003). Entwistle and Entwistle (1992) show that students are likely to study strategically depending on what is expected from the assessment. Consequently, it is reasonable to theorise that physics students have

significantly developed the skills that are most useful in these formal examinations such as knowledge recall and application of knowledge. Universities that aim to develop well-rounded and employable graduates are considering many facets of educational reform to target a wider range of skills needed in the workplace (Fry et al, 2003; Institute of Directors, 2007).

Momsen et al. (2013) conclude that examinations filled with questions testing recall may lead to students spending more time memorising and encourages a "surface-learning" approach. Students who view physics as a discipline of recall may develop a disconnected view of the subject, with consistency between ideas often coming from a learned cohesion that does not improve underlying understanding (Sikorski & Hammer, 2017). Moreover, Fernandez (2017) determined that students trying to develop understanding independently do not improve their overall performance and can become frustrated.

With attitudes shifting to educational reform, there have been several studies critiquing the type of conventional assessment that is frequently relied on in physics education. Similarly to the premise presented here, Darma et al (2018) find that typical exam questions tend to examine lower order thinking and suggest *authentic assessments* as an alternative (Pearce, 2016). These types of questions require students to create answers

Taxonomy	Definition
1 – Remember	Retrieving relevant knowledge from long-term memory.
2 – Understand	Determining the meaning of instructional messages, including oral, written, and graphic communication.
3 – Apply	Carrying out or using a procedure in a given situation.
4 – Analyse	Breaking material into its constituent parts and detecting how the parts relate to one another and to an overall structure or purpose.
5 – Evaluate	Making judgments based on criteria and standards.
6 – Create	Putting elements together to form a novel, coherent whole or make an original product.

**Figure 1** Revised Bloom's Taxonomy (Krathwohl et al, 2002)

by using critical thinking when given a realistic situation, and are seen to be a particularly good assessment method for Mathematics (Stobaugh 2013). The key here is the change of focus to the *creation* of an inventive answer, which leads students to access higher levels of cognition, according to Bloom's *Taxonomy* (1956) and to become better problem solvers. Skills that are well developed in students are likely to use levels of cognition that are frequently assessed, while underdeveloped skills are expected to lie in cognitive levels that are more rarely assessed. It follows that there could be a link between types of cognition and skill development that could be exploited, using examinations, since they are the dominant mode of assessment in many Physics degree programmes. With desirable skills in mind, such as those identified by Hanson and Overton (2010), this could give reason to consider modification of the styles of questions used in examinations and overall assessment techniques to better target nascent skills areas.

## Methodology

The aim of this research is to identify which skills are most developed by the process of unseen written examinations, in order to investigate how this commonly employed method of assessment influences the skill profile of physics graduates. Classification of exam questions was characterised by Bloom's Taxonomy to allow for links to be made between the tasks required and the cognitive skills needed to complete. Command words in

exam questions were identified and matched to cognitive levels in the Taxonomy based on the type of task they entail.

The analysis of exam papers allowed some investigation into the cognitive levels accessed by students in completing the examinations. There is also an established perception that formal examinations are fairer and efficient (Pugh, 2019). If this claim is true, it could be expected that all papers would be similar in cognitive difficulty from year to year in order to be deemed fair. This method of exam question analysis can also give some insight as to how effective examinations are, and if modifications are likely to make any impact. It may be possible to improve the ability of examinations to encourage a more rounded graduate skills profile by adjusting the phraseology of physics-based questions to improve higher order cognitive skills.

## Cognitive Interpretation of Skill Profile

The tool used to link exam questions and assessments to the cognitive order of tasks they test, and interpret the skill profile, was Krathwohl's Revised Bloom's Taxonomy (2002).

This is a hierarchy of cognitive skills, where each level displays a higher degree of cognition than the previous one, if Bloom's Taxonomy is considered to be hierarchical. It should be noted for the purpose of later discussions that a student may be capable of accessing one level without having perfected

Skill	Level in Bloom's Taxonomy
Oral Presentation	Understand, Create (2, 6)
Report Writing	Analyse, Evaluate, Create (4, 5, 6)
Team Working	Understand, Evaluate, Create (2, 5, 6)
Problem Solving	Apply, Analyse, Create (3, 4, 6)

**Figure 2** Categorisation of skills profile as defined by Bloom's Taxonomy

all levels below it this (e.g. a student may be able to *apply* an algorithmic calculation without having full *understanding* underlying concepts).

Krathwohl's (2002) definition of the Taxonomy (*Figure 1*) was used to categorise the key skill that is assessed through formal examinations, problem solving. When one acknowledges the traditional view of physics as a problem solving-based discipline, it may be surprising that physics graduates feel they have not sufficiently developed this skill (Hanson & Overton, 2010). Solving calculations accesses cognitive levels 3 and 4 (Apply and Analyse). However Momsen et al. (2013) find that physics students sometimes learn to mechanically solve problems without developing a deeper understanding that can be used in the workplace.

The graduate profile from the research by Hanson and Overton (2010) suggests that, based on this assignment of cognitive levels to skills, the most underdeveloped type of cognition in a physics graduate is *create*, with *evaluate*, *analyse* and *understand* also possibly underdeveloped, as indicated by the analysis of skills and the level of Bloom's taxonomy employed by these skills, as indicated in *Figure 2*.

Though the extent to which general skills are directly assessed may be evident (i.e. number of oral presentations required), this research investigates the frequency of testing of the cognition necessary for these skills (*Figure 2*) in the most common assessments – formal examinations. It is reasonable to expect a student who is well practised in accessing cognitive synthesis due to examinations will be more capable in skill areas that use this type of cognition than a student whose only experience of synthesis comes from the less

common, direct assessment methods. Thus, if examinations are a contributing factor to the consistent development of these skills this paper asks whether examination questions, as the dominant form of assessment in many Physics programmes, should test the evaluation and creation of ideas.

Authentic learning can provide a realistic scenario in an effort to get students to "create" their own answers, as opposed to algorithmic problem solving (Pearce, 2016). However, authentic assessments are not generally used in a closed-book exam environment. Consideration of alternative assessment methods is outside of the scope of this paper.

### Examination Analysis using Bloom's Taxonomy

This research reviews examination papers using Bloom's Taxonomy to identify where different levels of cognition occur. Questions in examinations will often have a measurable verb to set out the desired task, which indicates what kind of process is required to provide an answer. Through this relationship, these verbs can be categorised to correspond to a level of cognition in the Bloom's Taxonomy. Krathwohl et al (2002) provide some limited examples of measurable verbs that are representative of each level as this way of classification is common in educational research. The issue with this kind of categorisation is that it is subjective, depending on both the context of the question and the outlook of the reader. Some verbs fit indisputably into a level of Taxonomy. However most verbs are more difficult to definitively categorise when considered in isolation and out of context. Therefore it is important that the whole question, and not just the measurable verb, is considered when assigning a categorisation. In this paper, this measurable verb approach was used to attempt to match questions in

examinations to the levels of Bloom's Taxonomy that they require. The verb categorisation is based on research by Stanny (2016) that compared over 30 online compilations of such associations and details the frequency of each verb appearing as relating to each cognitive level in the papers. This paper also provides a modified table which necessitates a verb to have appeared in its assigned level in a least a third of the inspected group. Stanny's table was not used to classify verbs in this paper because in Stanny's work the papers analysed were across all subjects. As Bloom's Taxonomy is used across a multitude of subjects, using the Stanny's table may eliminate the instances that the command words are used in a mathematical sense, as is often the case for physics.

This basis reduces the subjective bias in classification but is somewhat limited as many commands appear in multiple levels due to context. This approach alone leaves the determination of the level a question requires, if it uses a command that appears in multiple levels of the Taxonomy, a subjective process. To address this, Krathwohl's (2002) definition of each level is used to refine the meaning of each verb (Figure 3). For example, the instruction "identify" appears in cognitive levels 1-4 in Stanny's paper (2016) but has been refined to include the focus of the instruction. When available, mark schemes were used in the analysis of sampled papers to help clarify exactly what type of process the command word entailed.

Four very common physics-specific meanings of command words have been added to words already listed in other levels in the original table:

- 1) Give [*definition*] – Level 1: Recall Existing Meaning: Give [*an example*] – Level 2: Understand Justification: It is common in physics to use "give" as an alternative to "define" when a formula or physical law is concerned. Example: "Give the heat engine statement of the 2<sup>nd</sup> Law of Thermodynamics."
- 2) Give [*calculation*] – Level 3: Apply Existing Meaning: Give [*an example*] – Level 2: Understand Justification: "Give"

is sometimes used in physics to mean "provide a value using a standard calculation". Example: "Give the Miller indices that describe each set of planes."

- 3) Estimate [*value*] – Level 3: Apply Existing Meanings: Estimate [*meaning*] – Level 2:
- 4) Understand, Estimate [*judgement*] – Level 5 Evaluate, Estimate [*original idea*] – Level 6: Create
- 5) Justification: Estimation in physics usually involves performing a standard calculation to provide an approximate value or mathematical answer, which is not meaning, judgement or original work. Example: "...Estimate where the next highest frequency of increased sensitivity will be found."
- 6) Added Meaning: Evaluate [*value*] – Level 3: Apply Existing Meanings: Evaluate [*relationships/changes*] – Level 4:
- 7) Analyse, Evaluate [*efficacy*] – Level 5: Evaluate, Evaluate [*original ideas*] – Level 6: Create Justification: Evaluate in a mathematical context often has the meaning "give a value for" and can be synonymous with "calculate" if the questions involves a numerical answer.

These additions made it less likely that a command word was wrongly categorised due to its uncommon usage in literacy-based disciplines. This alteration method is a useful way to adapt the Bloom's Taxonomy to cover other subjects and the technical language they may use.

Nonetheless, some instructions used in physics exam papers do not correspond to the words catalogued in the final table. These cases are instead either asked using phrases or subject specific verbs, such as "normalise" and "integrate". In these occurrences, the question is rephrased using one or more command words that preserve the original meaning and entail the same working out. For example, "integrate" can be asked equivalently as "calculate the integral". As "calculate" exists already in the categorisation table, the

Cognitive level	1. Remember	2. Understand	3. Apply	4. Analyse	5. Evaluate	6. Create
Measurable Verb (and explanation where applicable)			Calculate	Calculate (many formulae)		
	Draw/Sketch (standard diagram)					Draw/sketch (original diagram)
		Estimate (meaning)	Estimate (value)		Estimate (judgement)	Estimate (original idea)
		Explain (a term)	Explain (procedure)		Explain (judgement)	Explain (original idea)
			Evaluate (value)	Evaluate (relationship)	Evaluate (efficacy)	Evaluate (original idea)
	Give (definition)	Give (an example)	Give (standard calculation)			
	Indicate (from memory)	Indicate (from understanding)				
	Identify (definition)	Identify (meaning)	Identify (procedure)	Identify (relationship)		
		Illustrate (meaning)	Illustrate (procedure)	Illustrate (relationship)		
		Predict (meaning)	Predict (result of procedure)	Predict (relationship or change)		Predict (result of whole)
			Solve (standard formula)	Solve (many formulae or manipulation)		
	Write (down)		Write (procedure or standard formulae)			Write (original work)
			Determine (quantitative)		Determine (qualitative)	
	State					
	Recall					
	Label					
	Define					
		Discuss				
		Why (from understanding)				
		Demonstrate (meaning)				
			Show (that)			
			Demonstrate (procedure)			
				Deduce		
				Verify		
					Improve	

**Figure 3** Gates and Pugh's revised table of measurable verbs assigned to the cognitive levels of Bloom's Taxonomy (adapted from Stanny, 2016).

**A5.** (i) Explain the concept of thermal equilibrium incorporating the words *temperature* and *heat*.

(ii) What is the rate of heat flow along a lagged rod of thermal resistance  $2 \text{ KW}^{-1}$  when the ends of the rod are maintained at temperatures of  $100^\circ\text{C}$  and  $0^\circ\text{C}$  [4 marks]

Explain [procedure] – 3  
 "What is [X] when [Y]?" – Determine [X] when [Y] → Determine [quantitative] – 3

**Figure 4** Example of changing a question phrase into a measurable verb.

**A2.** A vertical circular loop of conducting wire of radius 2 cm rotates with angular speed 3 rad/s about the z axis (the vertical direction). The z axes passes through the centre of the loop and is in the plane of the loop at all times as the loop rotates. The loop is in a horizontal uniform magnetic field of  $25 \mu\text{T}$ . The resistance around the loop is  $50 \Omega$ .  
 Calculate the maximum current in the loop. [6 marks]

Calculate [many formulae/manipulation] – 4

**A3.** A long iron core solenoid of 6000 turns / m carries a current of 15 mA. At this current the relative permeability of the iron core is 1200.

(a) Calculate the magnitude of the magnetic flux density within the solenoid.

(b) If the iron core is removed, what current would be needed to produce the same magnetic flux density in the solenoid as before the iron core was removed? [4 marks]

Calculate [standard formulae] – 3  
 "What current would be needed?" → Calculate [standard formulae] – 3

**Figure 5** Example of the ambiguity in distinguishing a level 3 calculation from a level 4 calculation.

question can now be assigned a level based on this command word. Similarly, a question asked by a phrase in the format of "what is [X] when [Y]?" can rephrased as "determine [X] when [Y]" and categorised based on the measurable verb "determine" (Figure 4).

Although efforts were made to make this method of analysis more robust, there are still some inherent flaws with the process. One such issue is the decision of which level a task resides in when the command word appears in multiple levels. Though this is refined by the

definition of each cognitive level, as detailed above, does it necessarily mean that a task definitively belongs in just one cognitive level? For example, the word "calculate" can be assigned to "apply" if it involves just applying standard formulae, or assigned to "analyse" if it involves some manipulation or more complex formulae (see figures 5 and 6). Whether or not a calculation is "standard" or "complex" is subjective, though can be improved by some degree by considering factors such as the marks allocated for the task, the complexity of the mathematics in formulae used, and the

amount of formulae and manipulation necessary. Nonetheless, the distinction between levels 3 and 4 is not always conclusive (*Figure 5*).

Likewise there are instances in which there needs to be a subjective distinction between a task that could be said to be either recall or understand, and either evaluate or create. However, in the sampled papers, there is much less overlap between levels 2 and 3, and levels 4 and 5. Therefore it is sometimes useful to interpret results by combining the Taxonomy (similar to et al, 2014) into three complementary cognitive levels: Recall and Understand, Apply and Analyse, Evaluate and Create. These three revised levels, in the context of physics papers sampled, could be considered as: providing learned knowledge, using learned knowledge, and augmenting learned knowledge with original ideas and judgements.

Another drawback of this method of analysis is its inability to record tasks that are necessary for the resolution of a question but not explicitly detailed by a command word. This defect is most prevalent when considering that practically all questions include some element of concealed factual recall that is not stated explicitly. Nevertheless, the recall must be fulfilled before a student can access the actual command word and the cognitive level it represents. This can unintentionally make questions more difficult than questions written to access higher cognitive levels without a factual recall dependence (Ene & Ackerson, 2018) Physics undergraduate students are expected to have some quantity of formulae and definitions committed to memory.

However the collection of material that needs to be committed to memory is not always clearly defined at degree level, unlike at A-level and GCSE. It is reported by Fry et al (2003) that students need to memorise more formulae than ever before as syllabi have grown to include contemporary science, with only small amounts of original material taken out.

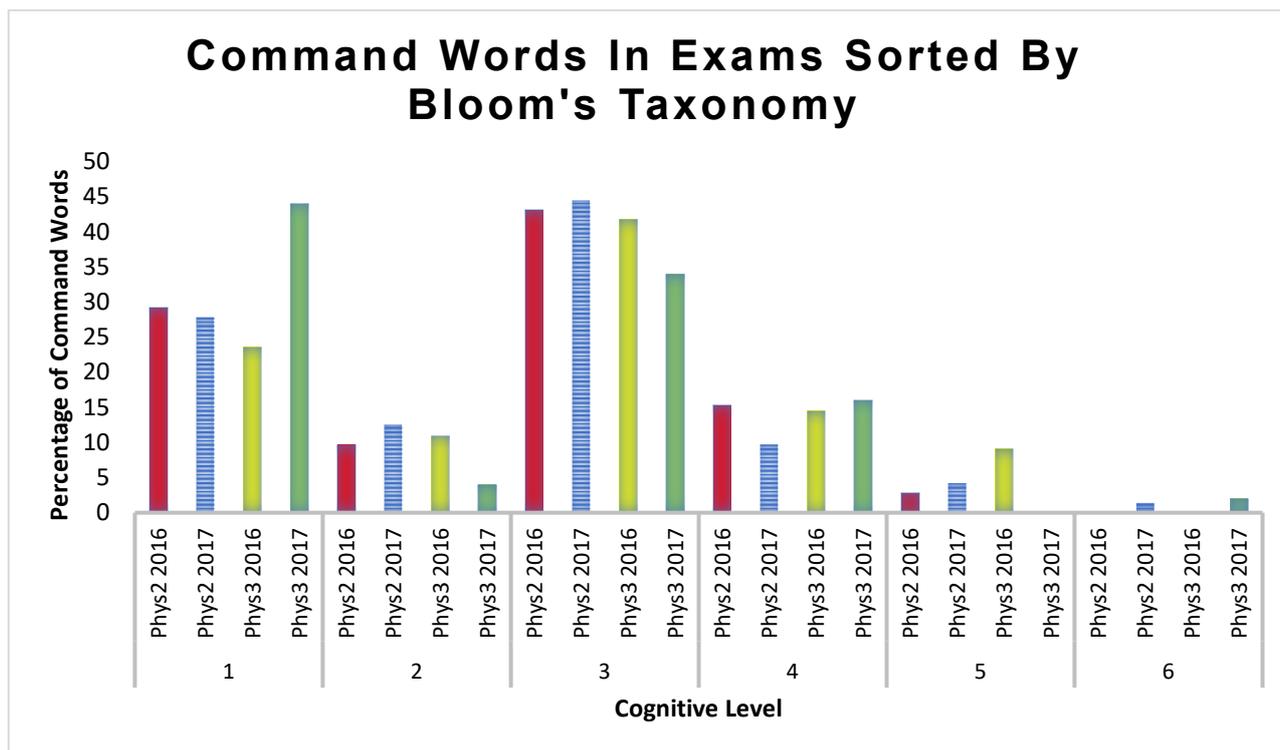
The nature of using command words to categorise means that the marks allocated for each command word, i.e. the weight each measurable verb carries, is not able to be captured. This can be advantageous in the

sense that the grading process at university is designed to include opportunities to give marks for methods that are not considered standard but are correct. However a by-product of this grading style is that the mark schemes do not commonly provide an exact mark for an individual step or task and often do not include smaller steps involved in calculation at all. Therefore an analysis based on assigning a precise mark to a specific task or word is erroneous by nature. It is also for this reason that it is not recommended to use a suggested solving sequence, as used by Ene & Ackerson (2018), to analyse these papers.

The use of command word frequency in each cognitive level as an analysis method is more robust but less informative. For the purpose understanding results, it is useful to generally assume that more time is necessary to perform tasks that reside in higher levels of the Taxonomy. This gives some idea of the weight of the command words without the uncertainty that is brought by estimating allocated marks.

## Findings and Discussion

Two modules were taken from a University core physics curriculum (named Physics 2 and Physics 3) and exam papers from 2016 and 2017 were analysed. These modules were chosen as they examine core physical theories including electromagnetism and thermodynamics, which are topics that have also been examined extensively by concept inventories (Lavery & Caballero, 2018). Formula sheets are not allowed during examination, nor is a definitive list of formulae that require memorisation provided. Even with this small sample of examination papers, *Figure 2* shows that these papers all have a similar cognitive profile. As exam papers should show similar trends as an assessment method, it is likely that the methodology is an effective tool of investigation for the sample group. It is worth noting that students sitting these examinations have a choice of questions and can achieve a pass grade with 40% correct, but in this study all questions on the paper were catalogued. Therefore the individual Taxonomy of the experience each student has of the exam can change slightly depending on their choice of questions and to what extent they were able to answer them.



**Figure 6** Cognitive profiles of sample exam papers.

The general distribution seems to indicate that examinations test recall (23-44%) and application (34-44%) more heavily than all other cognitive domains, with evaluation (0-9%) and creation (0-2%) tested the least (*Figure 6*). Although command words for analysis make up only 9-16% of total command words, it is worth noting that these words usually necessitate a multi-step problem that is expected to take a student longer than a level 3 “Apply” question and are therefore worth more marks (as noted in methodology). Also, the distinction between a level 3 calculation and a level 4 calculation is somewhat subjective. Questions examining understanding make up 4-13% of measurable verbs. However these are usually expected to take less time than calculations and are generally worth fewer marks. It is reasonable to assume that students working to the assessed task will focus most on areas that are worth the most exam marks and develop skills accordingly. With this fact considered, it can be deduced that the types of cognition that are well tested and encouraged by examinations are: Recall, Application and Analysis. It is possible that each module has a slightly different “characteristic cognitive profile”,

however a larger sample would be necessary to determine this effect.

Students may regard physics as a subject focused on calculations and “algorithmic problem-solving” (Momsen et al., 2013). The analysis based on Bloom’s Taxonomy shows that the cognitive levels where calculations generally lie, Application and Analysis or “using learned knowledge”, make up the majority of what is tested in these examinations. This may provide some explanation of the consensus from graduates (Hanson & Overton, 2010) that physics content is well developed. However it is obvious that evaluation and creation are not tested as commonly, although perhaps examinations are not the most suitable type of assessment to test these cognitive skills.

The Physics graduate skills profile, as defined by Hanson and Overton (2010) consists mostly of tasks that access the lesser tested cognitive levels: some Understand, and mostly Evaluate and Create (*Figure 2*). This implies that the dominance of formal examinations and the levels of cognition that they appear to test may be a contributing factor as to why these skills are underdeveloped in graduates. It could also

(e) Explain why conventional light microscopes have resolutions that are limited to around 300 nm, at very best. Explain briefly an experimental technique that would enable you to observe objects that are smaller than this. [4]

**Figure 7** Question B3)(e) in Physics 2 2016

be indicative that augmenting learned knowledge is underdeveloped in general and is comparable to Rustaman's (2017) proposition of underdevelopment of metacognition in physics graduates. Though there may be other assessment methods that are designed to directly address underdeveloped skills, research shows that including the tasks that need this cognition separately (e.g. only in lab reports or professional skills modules) is less effective for the overall cohesion of knowledge (Sikorski & Hammer, 2017). It can also serve to encourage students to adopt the epistemological expectation that core physics is a discipline of memorisation and mechanical calculation without need for original thought (Momsen et al. 2013).

### Recommendations for Further Research

It is possible that improving the breadth of cognitive levels tested in formal examinations is conducive to creating students with a more well-rounded skill profile. A broader cognitive profile could also positively influence the student's impression of physics as a whole, from a "mechanical" discipline to more percipient one, and improve their overall cohesion of knowledge.

There are several examples in the sampled papers where a question can be taken from targeting a low cognitive level to a higher/lesser-tested cognitive level by a small adjustment (Figure 7).

The first command, "Explain why...", refers to the procedure of determining the resolution of a microscope and provides end result. Using the methodology, this becomes explain [procedure] and resides in the third cognitive level: Apply. The majority of "Explain" questions found in the sampled papers are similar and ask for an explanation of either a term (level 2 – understand) or a procedure. However the second instance of this word shows how a very slight change in context can

change the type of thinking a question requires. This example actually asks the student to use their judgement to explain how to overcome a given failing, accessing the cognitive level of evaluation. A student must evaluate the efficacy of an alternative process to determine qualitatively if this method is better than a conventional light microscope. This could also be considered a form of "authentic assessment" as it gives information on a real situation and asks the student to form an answer based on this scenario (Sugiarti et al., 2017). These types of adjustments could be engineered by considering the table of measurable verbs (Figure 3) and altering the task each command word describes. For example, a question that originally asks a student to sketch a standard or memorised diagram (recall) could be improved by asking them to sketch the diagram under some kind of constraint, thus making their result an original idea (sketch [original diagram] is in *create*).

There are several ways in which this type of exam analysis can be developed further. For example, as is evident from the qualitative investigation into formulae recall, some equations are more heavily weighted than others. This insight could be used to improve the method by including a way of cataloguing marks gained from a correct formula recall against possible marks lost if the formula is not recalled successfully, and investigate if this has an effect on achievement. With a larger sample base, it would be possible to investigate the degree to which the cognitive profile and formulae recall affect grade distribution. This would improve the predictive power of this method and in turn provide a possible moderation structure to maximise the objectivity of examinations from year to year, even before they are implemented as assessments. In conjunction with this, it is conceivable to adapt this methodology to interpret other forms of assessment to give

perspective on a degree course as whole or adapt it to fit other cognate disciplines.

It is possible that expanding the breadth of cognition that is assessed in formal examinations may foster an improvement in higher-order thinking, cohesion of knowledge and sequentially skill development, and this is an area that should be considered further. Any modification to move towards this change to examinations or the overall assessment regime should also consider memorisation as barrier to desired tested cognition.

In terms of assessing the skill profile of graduates, this research seems to be able to link cognition necessary for a skill and frequency of assessment to give some idea of how developed it might be. However, the actual development of skill is influenced by many other factors, such as teaching styles, competence in the affective domain (particularly with oral skills) and learning objectives. With consideration of factors such as these, there is the potential to use this method to reverse-engineer assessments based on desired skills and learning objectives. Furthermore, determination of a characteristic cognitive profile for each module could help educators to understand the skill set that is encouraged at different stages throughout a programme.

## Conclusion

A method for analysing examination papers in Physics based on the assignment of measurable verbs to levels of cognition in Bloom's Taxonomy was developed and applied. The most commonly used verbs in examination questions were in the recall and apply categories of cognition.

Based on parity between the cognition used in the skill profile and the frequency of tested cognition in exam papers, it may be concluded that reliance on formal examinations could be a contributing factor in the underdevelopment of wider skills in students. Specifically, examinations are excellent at testing application of knowledge but rarely encourage the utilisation of original thought.

## References

- Bates, S.P., Galloway, R.K., Riise, J. & Homer, D. (2014) *Assessing the Quality of a Student-Generated Question Repository*. Physical Review Special Topics-Physics Education Research 10, no. 2 (July 18, 2014): 020105. DOI: 10.1103/PhysRevSTPER.10.020105.
- Bloom, B.S. (Ed.). Engelhart, M.D., Furst, E.J., Hill, W.H. & Krathwohl, D.R. (1956). *Taxonomy of Educational Objectives, Handbook I: The Cognitive Domain*. New York: David McKay Co Inc.
- Clark, D. (2005) *Bloom's Taxonomy of Learning Domains: The Affective Domain*. <http://www.nwlink.com/~donclark/hrd/bloom.html> (Accessed 29/04/2021)
- Darma, I.K., Candiasa, I.M., Sadia, I.W. & Dantes, N. (2018) *The Effect of Problem Based Learning Model and Authentic Assessment on Mathematical Problem Solving Ability by Using Numeric Ability as the Covariable*. Journal of Physics: Conference Series 1040 June: 012035. DOI:10.1088/1742-6596/1040/1/012035
- Ene, E. & Ackerson, B.J. (2018) *Assessing Learning in Small Sized Physics Courses*. Physical Review Physics Education Research 14, no. 1, January: UNSP 010102. DOI: 10.1103/PhysRevPhysEducRes.14.010102
- Entwistle, N. & Entwistle, A. (1992) *Experiences of Understanding in Revising for Degree Examinations*, Volume 2, Issue 1, March, pp. 1-22. DOI: 10.1016/0959-4752(92)90002-4
- Fernandez, F.B. (2017) *Action Research in the Physics Classroom: The Impact of Authentic, Inquiry Based Learning or Instruction on the Learning of Thermal Physics*. Asia-Pacific Science Education 3, no. 1, December. DOI: 10.1186/s41029-017-0014-z
- Fry, H., Ketteridge, S. & Marshall, S. (2003) *A Handbook For Teaching & Learning in Higher Education*, Second Edition.

Hanson, S., & Overton, T. (2010) *Skills required by new physics graduates and their development in degree programmes*. Higher Education Academy UK Physical Sciences Centre.

Institute of Directors (2007) *Institute of Directors skills briefing – December, Graduates' employability skills*. 116 Pall Mall, London SW1Y 5ED.

Institute of Physics (2010), The Physics Degree: Graduate Skills Base and the Core of Physics  
<https://www.abdn.ac.uk/careers/resources/documents/4581.pdf> (accessed 23/5/2021)

Krathwohl, D.R. (2002) *A Revision of Bloom's Taxonomy: An Overview*. *Theory Into Practice* 41, no. 4, pp.212–18. DOI: 10.1207/s15430421tip4104\_2

Laverty, J.T. & Caballero, M.D. (2018) *Analysis of the Most Common Concept Inventories in Physics: What Are We Assessing?* *Physical Review Physics Education Research* 14, no. 1 April: 010123. DOI: 10.1103/PhysRevPhysEducRes.14.010123.

Momsen, J., Offerdahl, E., Kryjevskaja, 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* vol. 12, 2, pp. 239-49. DOI: 10.1187/cbe.12-08-0130

Pearce, S. (2016) *Authentic Learning: What, Why and How? E-Teaching: Management Strategies for the Classroom*. Australian Council for Educational Leaders, April, Vol. 10,  
[http://www.aceel.org.au/aceel/ACEL\\_docs/Publications/e-Teaching/2016/e-Teaching\\_2016\\_10.pdf](http://www.aceel.org.au/aceel/ACEL_docs/Publications/e-Teaching/2016/e-Teaching_2016_10.pdf) (Accessed 29/04/2021)

Pugh (2019) *Rethinking University Assessment by Learning from Secondary Education*, University of Leeds,  
<https://teachingexcellence.leeds.ac.uk/teaching-enhancement-projects-current/samantha-pugh-school-of-physics-and-astronomy/>

Rustaman, N. Y. (2017) *Assessment in Science Education*. *Journal of Physics: Conference Series* 895 (September 2017) 012141. DOI: 10.1088/1742-6596/895/1/012141

Stanny, C.J. (2016) *Re-evaluating Bloom's Taxonomy: What Measurable Verbs Can and Cannot Say about Student Learning*. *Education Sciences* 6, December. DOI: 10.3390/educsci6040037

Sikorski, T-R. & Hammer, D. (2017) *Looking for Coherence in Science Curriculum*. *Science Education* 101, no. 6, November, pp. 929–43. DOI: 10.1002/sce.21299

Stobaugh, R. (2013) *Scenarios and Real World Applications*. Assessing critical thinking in middle and high schools: Meeting the common core, Chapter 4, Routledge, ISBN 9781315853451.

Sugiarti, T., Kaniawati, I. & Aviyanti, L. (2017). *Development of Assessment Instrument of Critical Thinking in Physics at Senior High School*. *Journal of Physics: Conference Series* 812, February, 012018. DOI: 10.1088/1742-6596/812/1/012018

The Quality Assurance Agency (2017) *The QAA Benchmark Statement for Physics, Astronomy and Astrophysics*,  
[https://www.qaa.ac.uk/docs/qaa/subject-benchmark-statements/subject-benchmark-statement-physics-astronomy-and-astrophysics.pdf?sfvrsn=eff3c881\\_4](https://www.qaa.ac.uk/docs/qaa/subject-benchmark-statements/subject-benchmark-statement-physics-astronomy-and-astrophysics.pdf?sfvrsn=eff3c881_4) (Accessed 24/04/2021).