



Rachel Archer and
Simon Bates
School of Physics and
Astronomy
University of Edinburgh
Edinburgh
ED
r.s.k.archer@sms.ed.ac.uk
s.p.bates@ed.ac.uk

Asking the right questions: Developing diagnostic tests in undergraduate physics

Abstract

Being able to discover students' conceptions and more importantly alternate- and misconceptions about a topic is vital in order to be able to assess and thus be able to improve student learning. It is well known that this can be achieved via the use of well-designed diagnostic tests, a widely used example of which is the Force Concept Inventory. Creating the right questions in order to form a reliable diagnostic test can be a lengthy and complicated process. This article reports work on a Development Project funded in 2008 to develop such a test for introductory Quantum Mechanics courses in both physics and chemistry. We present details of our methodology, which involves augmenting a 'standard' multiple-choice question set with free-response boxes to determine the reasons for a student choosing a particular answer, and a self-assessment of their level of confidence in their choice. The responses from piloting this initial test in different institutions are used to inform the subsequent refinement of the test, as well as assessing the reliability and validity of the questions. We highlight examples of misconceptions that have been found during the development of the diagnostic tests.

Utility of Diagnostic Tests

Diagnostic tests have been used in a range of different subjects in order to gain a fuller understanding of students' grasp of key concepts. Misconceptions can occur 'when prior knowledge and belief are in conflict with scientific knowledge,'¹ and can cause many problems for students during a course and throughout their degree. It thus seems essential for instructors to uncover these misconceptions since there is strong evidence that these need to be taken into account in order to improve the efficiency of instruction.² The Force Concept Inventory² is an example of a diagnostic test in classical mechanics that has been shown to uncover many misconceptions that students hold about the subject. This test is a set of multiple-choice questions and is generally given to the students both pre and post instruction. Richard Hake³ used this diagnostic test and surveyed approximately six thousand students in high schools, colleges and universities in the US. The test was used as a quantitative measure of learning gains from different types of instruction and so was given to the students both pre and post instruction and from this their percentage gains were able to be calculated.³ It can be interpreted from these results that the test consistently highlighted certain misconceptions, regardless of prior learning of the subject or instructional style. Once a test has been constructed it is important to see how valid and reliable it is. If a test is reliable it is said to be 'consistent within itself and consistent across time.'⁴ A test will be valid 'if the skills or knowledge it measures are directly relevant to the stated domain of the test.'⁴ There are ways of testing whether a test is reliable and valid and more information can be found in reference 4.

Styles of Questions

When designing a diagnostic test that is going to be an effective tool in uncovering any misconceptions to which the students may be subject, it is important to ascertain the origin of these misconceptions. This information can then feed forward into redesigning and improving instruction.

We have employed three different approaches in designing our tests, which consist of three distinct styles of multiple-choice questions. One of the reasons for providing the multiple-choice questions in different styles and media formats, i.e. on paper or online, is so that instructors will be able to utilise the one that suits their needs best. The first style is the standard multiple-choice question. This consists of the question and the selection of possible answers from which the students select. This type of question is widely-used but may not yield information as to why the students are under particular misconceptions. However, it is very easy to deploy online as well as on paper and it is the simplest to analyse.

Diagnostic tests have been used in a range of different subjects in order to gain a fuller understanding of students' grasp of key concepts.

The second type of multiple-choice question is the free response with confidence level question. This format consists of the standard multiple-choice question followed by a free response text box for the students to write down why they think their selected answer is correct, then a selection of confidence levels for them to choose how confident they feel with their reasoning. An example from our diagnostic test is shown in Figure 1.

This form of question provides the most information about why students may be under a certain misconception. It can also indicate whether the question tests what you think it should test from a student's perspective: this is one aspect of assessing question validity. It does take far longer to analyse than the standard multiple-choice question but is still relatively simple to deploy on paper and online. The responses from this style of test can then feed forward into the creation of the third type of multiple-choice question.

Two-tier multiple-choice questions⁵ are the third style of questions used in diagnostic tests. These can be most effectively created from the information provided from the free response text boxes in the previous style of question. This format provides almost as much information as the previous one as to why students may hold certain misconceptions, but it is far simpler to analyse. An example is shown in Figure 2.

The main potential drawback with this style of question is the fact that students taking the test may use the selection of reasons as to why the answers may be correct in order to assist them to choose their answer. A way of preventing this problem is by deploying this style of question online, separating the answers from the reasons why and preventing students from switching back and forth between the two selections. This would not be possible to prevent when deploying the test on paper, which may thus limit its use.

Designing Questions

In order to find the right questions to construct a well-designed diagnostic test the first step is to find the key concepts on which the test should be based. Starting with a list of the core topics covered in a course it is then possible to create a concept list. Once the key concepts have been determined the questions can then be developed based on these. For the development of our tests, we surveyed a variety of syllabi in both physics and chemistry instruction in quantum mechanics at different institutions. When the question has been created the selection of answers from which the student will choose needs to be carefully designed. Each selection should be both

IF YOU DO NOT KNOW THE ANSWER TO A QUESTION PLEASE LEAVE IT BLANK 8

8 Tunneling

The total energy of an electron after it tunnels through a potential energy barrier is...

A ...greater than its energy before tunneling

B ...equal to its energy before tunneling

C ...less than its energy before tunneling

Please state your reasons why you think your selected answer is correct

How confident are you in your reasoning?

Very sure Fairly sure A little unsure Very unsure

Figure 1: An example of a free response with confidence levels multiple-choice question, which was taken from the pilot study for the introductory Quantum Mechanics test.

plausible and ideally represent a certain misconception a student may hold. It has been suggested that 'The quality of a multiple-choice question, could be said to be based upon the quality of the distracters, not the quality of the question.'⁶

One measure of the validity of the questions is to employ the use of 'expert validity'. This consists of expert opinions on the face validity (seeing whether the concepts tested are related to the subject) and content validity (coverage of the subject matter) of the items on the test.⁴

Once the questions have been fashioned it is then possible to group them into three general categories: 'Recall', 'Interpret' and 'Apply'.⁷ Where 'Recall' questions simply ask the student to recall key facts and definitions, 'Interpret' questions require

Tunneling

The total energy of an electron after it tunnels through a potential energy barrier is...

A. ...greater that its energy before tunneling

B. ...equal to its energy before tunneling

C. ...less than its energy before tunneling

What is the reason for your selected answer?

1. Due to the conservation of energy the electron's energy remains constant
2. Energy is gained from the potential energy barrier
3. The energy of the electron exponentially decays within the barrier
4. Energy is required to tunnel through the barrier and is then dissipated

Figure 2: An example of a two-tier style question taken from the revised question shown in Figure 1.

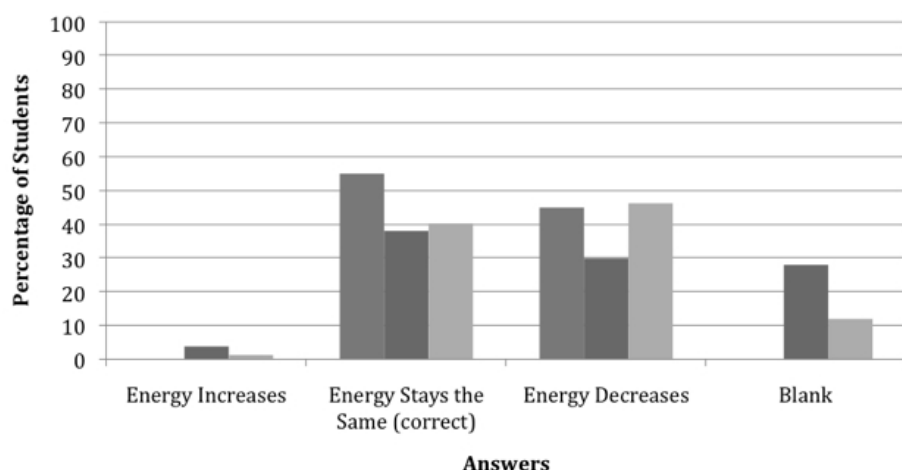


Figure 3: Results from Quantum Tunneling Question (Figure 1) for the three institutions post instruction

on which questions were based for the focus groups in order to gain clarity. With the information gathered from these groups some of the questions will be revised and others removed from the test.

Misconceptions

The Quantum Mechanics diagnostic test uncovered several misconceptions, one of which is that 'energy decreases when an electron tunnels through a potential energy barrier'. This misconception has also been discovered from the 'Quantum Mechanical Conceptual Survey' at the University of Colorado.⁸ It has been prominent throughout all three institutions as well as persisting post-instruction at the University of Edinburgh. This is illustrated in the results obtained

students to extrapolate 'already learned material in a qualitative fashion', and 'Apply' questions require both extrapolation and 'numerical manipulation'.⁷ Ideally when designing a conceptual diagnostic test, it would be best to have most of the questions residing in the 'Interpret' category.

Piloting the Test

At the University of Edinburgh we have devised two diagnostic tests for introductory Quantum Mechanics in both physics and chemistry. They have both employed the same method of design as explained earlier, where the free response text boxes in the pilot test were used to feed forward into the creation of a two-tier version of the test. This methodology has also been used in the creation of other diagnostic tests at the University of Edinburgh.

The introductory Quantum Mechanics physics pilot test was deployed to second-year Edinburgh University physics students both pre and post instruction. By delivering the test both pre and post it was then possible to ascertain any misconceptions that improve with teaching and those that persist throughout and require further attention. This test has also been deployed to second-year University of Glasgow and University of St Andrews physics students post instruction. From the three institutions there were 134 students who took the test post-instruction. Analysis of these results is on-going and the test is currently being revised to create a two-tier edition.

Revisions of the Tests

Whilst analysing these results it has been seen that there has been very little difference between the spread of results between the various institutions, implying that the different institutions are relatively homogeneous in terms of ability post-instruction on introductory Quantum Mechanics courses and so implying that this diagnostic test is then widely applicable. However, there have been a couple of 'rogue' questions discovered. These have been, for example, where the majority of students have left a question blank. In order to find out more information about students' understanding of these questions and concepts several focus groups have been held. Some of the preliminary analysis directly informed the topics

from the selected answers to this question (Figure 1) as shown in the graph (Figure 3). Another misconception that was uncovered involved the students confusing the spacing of energy levels of an infinite potential square well with that of a hydrogen atom. This misconception was also prominent throughout all three institutions and did persist post-instruction at the University of Edinburgh.

Preliminary Conclusions

Diagnostic tests have been shown to be exceptionally useful in exposing students' conceptions of a subject. However, it is a lengthy and time-consuming process in order to develop them. Although the tests will still need further revisions in the future, they are still able to demonstrate to instructors what concepts students are and are not understanding in their course. The results from these tests will hopefully be carried forward into any revisions of the courses in the future. The project will be completed by Autumn 2009 and so the quantum mechanics diagnostic, which is currently under revision, should be available mid-Sept. It will be made available from the PSC Development Project website for this project:

http://www.heacademy.ac.uk/physsci/projects/detail/development_projects_2008/bates

Acknowledgements

We would like to thank all staff and students at the institutions who contributed to the project, as well as both the Physical Sciences Subject Centre and the Principal's Teaching Award Scheme at the University of Edinburgh for funding.

References

- Garcia Quijas, P. C. Arevalo Aguilar, L. M. (2007). Overcoming misconceptions in quantum mechanics with the time evolution operator. *Eur. J. Phys.* **28** (2007) 147–159.
- Hestenes, D. Wells, M. Swackhamer, G. (1992). The Force Concept Inventory. *The Physics Teacher*, **30**, March 1992, 141-158

3. Hake, R. R. (1998). Interactive engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *Am. J. Phys.*, 66, 1.
4. Ding, L. Chabay, R. Sherwood, B. Beichner, R. (2006). Evaluating an electricity and magnetism assessment tool: Brief electricity and magnetism assessment. *Phys. Rev ST Phys. Educ. Res.* 2 010105.
5. Treagust, D. F. Chandrasegaran, A. L. (2007). The Taiwan national science concept learning study in an international perspective. *International Journal of Science Education*, 29, (4), 391-403.
6. Davies, P. (2002). 'There's no confidence in Multiple-Choice Testing, in: *Proceedings of 6th CAA Conference*, Loughborough: Loughborough University.
7. Aubrecht, G. J. Aubrecht, J. D. (1982). Constructing objective tests. *Am. J. Phys.*, 51, 614.
8. McKagan, S. B. Perkins, K. K. Wieman, C. E. (2008). A Deeper Look at Student Learning of Quantum Mechanics: the Case of Tunneling. *Phys. Rev. ST Phys. Educ. Res.* 4, 020103.

The Quantum Mechanics diagnostic test uncovered several misconceptions, one of which is that 'energy decreases when an electron tunnels through a potential energy barrier'.