

RESEARCH DIRECTIONS

A Measurement of Quantum Fears

J.R. Chapman, S.O. Hunt & N.R. Walet

School of Physics and Astronomy, University of Manchester, Manchester M13 9PL UK

Abstract

There are anecdotal indications that students avoid questions involving time-dependent exam problems in quantum mechanics. To obtain real evidence, a diagnostic questionnaire has been created and administered to third-year students. It measures students' understanding of quantum mechanics in general, with an emphasis on misconceptions and threshold concepts that may block a deeper understanding of quantum mechanics, especially of time-dependent aspects. The questionnaire consists of two parts, a self-evaluation section followed by a conceptual survey. Analysis of the results of this questionnaire does indeed reveal areas of weakness in student understanding of time-dependence as well as of other fundamental quantum mechanical concepts. The questionnaire has been revised in light of the analysis with the aim of improving the understanding of student difficulties as well as the reliability of the questionnaire itself.

Keywords: quantum mechanics teaching, diagnostic tools, time-dependence

Introduction

Quantum mechanics is undoubtedly one of the most challenging theories in modern physics. Whilst it allows us to describe phenomena at the atomic scale, it is probably also one of the most philosophically challenging theories with which our students come into contact. Since it has a wide applicability, it is important for physics students to develop a thorough understanding of the subject. However, as a subject of study quantum mechanics presents a number of issues, encapsulated in Richard Feynman's oft-repeated remark: "I think I can safely say that nobody understands quantum mechanics". (Feynman 1990a) Whilst the

Corresponding author:

Professor Niels Walet, School of Physics and Astronomy,
University of Manchester, Manchester M13 9PL, UK
Email: Niels.Walet@manchester.ac.uk

mathematical and technical content of a *second* quantum mechanics course – the subject we concentrate on – is probably less challenging than in other courses (e.g. electromagnetism), it contains many difficult concepts that require students to make a departure from their common-sense understanding of the world. Even though in a second course we usually expect that students have moved beyond crude analogies to help them understand principles such as wave-particle duality or quantum-mechanical tunnelling, the evidence is that students still struggle with many of the concepts. We would like to know whether students can deal with the problem in the way Feynman (1990b) describes it: “They must accept Nature as She is – absurd.”

There is anecdotal evidence that students studying quantum mechanics have a tendency to avoid examination questions involving time-dependence. Whilst within an examination with a choice of questions (as is the case in the UK) it is not unusual for students to attempt to avoid those relating to what they perceive as a difficult section of their courses, in quantum mechanics examinations the questions on time-dependence are often superficially less technically difficult than other problems on the paper. Of course, time-dependence and superposition of waves play a central role in developing a thorough understanding of all wave phenomena, not only in quantum mechanics. In this article we present a tool that can be used to assess students' conceptual comprehension of some of the core ideas of quantum mechanics, whilst identifying any specific barriers to understanding that may prevent students from being able to approach time-dependence with confidence. We investigate this instrument with a group of third year students.

The term ‘conceptual understanding’ refers to the ability to relate a purely mathematical description to a real-world result, i.e. an understanding of the contextual meaning of the mathematics, not just the ability to calculate a solution. Conceptual understanding is separate from academic success (whether in examinations or coursework) as it corresponds to a complete and thorough understanding of the subject rather than to the ability to complete situational problems. This depth of approach is not universally accepted, especially in quantum mechanics. Many physicists argue that quantum mechanics is best left as a tool for making predictions; that students should not try to understand the inner workings of quantum systems, as there are many interpretations that fit within the mathematical framework. However, in terms of learning the subject this is a less than ideal solution, as for many students a mathematical framework without a contextual scaffold is meaningless Singh *et al.* (2006).

With this in mind we decided that the best course of action was to develop a questionnaire to assess student understanding of quantum mechanics: the Quantum Mechanics Diagnostic Questionnaire (QMDQ). It is designed to be taken both at the start and end of term. We report here a test at the start and end of the third year course Applications of Quantum Mechanics. Questionnaires such as this are particularly useful as they can be used to assess a large number of students with relative ease, and they are particularly conducive to statistical analysis. The technical content of the questionnaire is at the second year level, looking to measure student competence based around the key learning outcomes of this course. The non-technical content is aimed at assessing the students' confidence in their ability. We assessed the students as soon as the questionnaire was ready (a few weeks into the semester). In the standard manner used for such questionnaires we also added an end-of-course questionnaire. However, it should be noted that for this project, whilst the 2nd QMDQ was distributed at the end of term, only a few student responses were received.

Initial project outline

At the University of Manchester we have two different groups of quantum courses, which for most students are open to choice: a ‘fast stream’ for more theoretically inclined students, and a separate set of courses for all other students. The students we have investigated in this study were in the third year of their degree, taking their second QM course (Applications of Quantum Mechanics or AQM) in the normal stream. In the 2012/13 academic year this group consisted of 135 students. All students take the second year Introduction to Quantum Mechanics (or IQM) course.

The aims of this project were to create and develop a diagnostic tool that can be used to:

- test students' understanding of quantum mechanics;
- substantiate evidence of students' under-performance in answering time-dependent quantum mechanics questions;
- test student opinion of, and confidence in, quantum mechanics;
- identify areas of poor performance in time-dependence, and possible causes.

The conceptual questionnaire was chosen as the key instrument for its capacity to study a large group and provide data suited to statistical analysis. Upon identification of the areas where students

en-counter difficulty, thought can be given to the content and theme of such student interviews.

Within PER there are two main categories of student assessment: concept inventories that look to characterise student performance in a single narrowly defined concept; and surveys that look at a broad range of topics within a subject (see Edinburgh archive of conceptual tests in physics, <http://bit.ly/ConceptTests> (accessed 14/7/2013)). This project, whilst having a focus on time-dependence, takes a more holistic approach to assessment and thus falls into the latter category. Modern and classical test theory states that "In test construction, a general goal is to arrive at a test of minimum length that will yield scores with the necessary degree of reliability and validity for the intended uses" (Adams & Wieman 2011); this elicits a unanimous consensus within the PER community (Lindell *et al.* 2006). In our context reliability is how well the survey discriminates between candidates, and whether results can be seen to be consistent. Validity refers to whether the survey covers sufficient course material, and whether the interpretation of results can be said to be meaningful (Wutti-prom *et al.* 2009). The processes that we have used to ensure the QMDQ is both reliable and valid are discussed below.

On the topic of validity, many researchers have expressed the need for clarity in survey questions, that wherever possible they should avoid excessive use of jargon (Wutti-prom *et al.* 2009). In particular, it has been noted that many students misread questions, "overlooking the critical role of 'little words' such as prepositions that determine meaning" (see Edinburgh archive of conceptual tests in physics, <http://bit.ly/ConceptTests> (accessed 14/7/2013)). More importantly consideration must be given to a student's interpretation of the question (Maloney *et al.* 2001), i.e. to whether a student is answering the question we think we are asking. If not, then the answer they give, regardless of whether it is right or wrong, is meaningless. It has also been noted that difficulties arise in constructing questions to which experts do not have obscure objections (Falk 2007). Addressing these concerns often requires the addition of conditions, or obtuse phraseology, which complicates the question. So there must be a trade-off between the theoretical accuracy of a question and how well it is understood by the average student. It has also been suggested that the incorrect answers given by students are more useful than the correct answers when determining their overall understanding of a topic. An incorrect answer represents the point at which the students' conceptual understanding starts to fail. Through the use of well-designed distractors (either incorrect multiple choice options or false statements in true-false questions) that favour

students' preconceptions, it is possible to highlight areas where student learning has been superficial. We have used this technique to examine some of the more fundamental and important concepts within the IQM course, using distractors based on the common misconceptions discussed in the next section.

We looked at the Quantum Physics Conceptual Survey (QPCS) (Sharma 2012) This QPCS was used to test conceptual understanding of first and second year undergraduates at the University of Sydney. The QPCS was created by Sydney University Physics Education Research group (Styer 1996). This survey covers basic quantum physics concepts, such as wave-particle duality, photoelectric effect, de Broglie wavelength, double slit interference and the uncertainty principle. All 25 questions are multiple-choice, with questions often sharing the same answer set or initial set-up. This survey was found to be effective at highlighting gaps in student understanding, but is probably not relevant for students at the more advanced level we are considering.

The conceptual surveys reviewed cover the base level of knowledge required for introductory physics and are suitable for first-year (UK) undergraduate students. The AQM we are investigating are at a more advanced level in their understanding, therefore conceptual probing is required at a higher level. Through the study of core modules in previous years it was determined that two courses were particularly key in building understanding of quantum mechanical and time-dependent systems. These courses are Introduction to Quantum Mechanics mentioned before and Mathematics of Waves and Fields, both of which are courses in semester 3 (first semester in second year of study).

To give an idea of the expected level, the learning outcomes for IQM include the ability to:

- understand how quantum states are described by wave functions;
- solve the Schrödinger equation and describe the properties of a particle in simple potential wells;
- solve one-dimensional problems involving transmission, reflection and tunnelling of quantum probability amplitudes;
- demonstrate an understanding of the significance of operators and eigenvalue problems in quantum mechanics;
- demonstrate an understanding of angular momentum in quantum mechanics.

Learning outcomes from the mathematics course, include the ability to:

- solve partial differential equations using the method of separation of variables;
- define the term "orthogonality" as applied to functions, and recognise sets of orthogonal functions that are important in physics (e.g. trigonometric functions and complex exponentials on appropriate intervals, Legendre polynomials, and spherical harmonics);
- solve eigenvalue problems (differential equations subject to boundary conditions) either in terms of standard functions or as power series;
- use sets of eigenfunctions as basis functions.

This knowledge of students' course history allows the creation of a diagnostic tool focused on content that has already been taught to students. Any misconceptions highlighted will therefore be of concern, as they have not been corrected in either of the completed courses. Analysis of the results of the survey could then show whether students find time-dependent questions harder than other areas and also locate common misunderstandings or misconceptions.

Planning the Quantum Mechanics Diagnostic Questionnaire

The diagnostic survey we built for this project is split into two sections. The first part tests students' preferences and confidence in a series of self-evaluation questions; the second part covers conceptual questions which test student understanding. The self-evaluation questions were included to investigate any link between confidence, learning style preference and performance on questions related to quantum physics and time-dependence.

In terms of a minimum requirement for the understanding of quantum mechanics, the Copenhagen interpretation is universally used as the baseline in both teaching and quantum mechanics PER (<http://bit.ly/ConceptTests>, Wuttirom *et al.* 2009, Falk 2007, Cataloglu & Robinett 2002, Singh 2001, Cramer 1986). The minimum conceptual understanding of this interpretation can be summarised as follows (Aubrecht and Aubrecht 1983):

- Heisenberg's uncertainty principle, including the concept of wave-particle duality;
- Born's statistical interpretation, describing a particle's behaviour through the wave function of Ψ and probability $P = \Psi^* \Psi$;
- Bohr's concept of complementarity and the intrinsic nature of uncertainty. Heisenberg's

state vector and collapse of the wavefunction;

- Heisenberg's Positivism, i.e. the capacity to verify results with experimental measurement.

Of these key concepts, the most common misconceptions are found to be focused within a few specific regions. Whilst students demonstrated an understanding of wave-particle duality, the concept of complementarity (specifically simultaneous particle and wave behaviour) is continually contradicted with descriptions of particles with 'wave parts & classical parts' (Falk 2007). Understanding the nuances of the wave-function also seems to cause students considerable trouble. Whilst many demonstrate proficiency in calculations, many misconceptions revolve around the physical interpretation of a wavefunction. Common mistakes are: associating the amplitude of the wave function with energy rather than probability, and stating that particles lose energy in tunnelling (Cataloglu & Robinett 2002).

In the conceptual part, items measuring student ability on time-dependent questions and other related areas are included. A mixture of interpretative and non-interpretative questions is included to evaluate both the level of student ability and conceptual understanding. Interpretative questions require understanding to answer correctly, and are often based on the Copenhagen interpretation of quantum mechanics (Wuttirom *et al.* 2009). Non-interpretative questions are more straightforward and rely on understanding or knowledge of the system in question. For students to score highly on the questionnaire they must recall the learning outcomes from previous courses and show an understanding of the meaning behind the concepts presented.

Question topics for the QMDQ

From this research and through discussions with experienced professors, it was concluded that the diagnostic questionnaire should contain approximately 35 questions split into the following groups:

- self-evaluation at the level of the IQM course in confidence and understanding;
- general quantum physics questions;
- mathematical questions;
- questions related to time-dependence.

The questions need to be a mixture of interpretative and non-interpretative and to cover a range of topics, including time-dependence, probability, eigenfunctions and concepts in

quantum mechanics. The learning outcomes from previous courses are not to be directly tested, but will be used to inform the difficulty of questions asked and the content covered.

Upon satisfactory design of this diagnostic questionnaire it was given to students for completion, via a link provided through the student email system. The results from this questionnaire were then analysed using the statistical methods outlined earlier. From the critical analysis of the QMDQ, improvements can be made to ensure that the final version of the QMDQ is a reliable and valid diagnostic resource that can be used by other groups in further investigations. The revised questionnaire could then be given to the same students when they complete their third year quantum mechanics course to look for improvements or patterns between the two sets of results. Testing students at different points of their education can also reveal the effect of new instruction.

Quality and validity

When designing questions for the questionnaire attention was paid to the quality of the questions such that they would be suitable to present to students. It can be difficult to write a good question that is clear, balanced and probing. When drafting the questionnaire some principles were followed (Lindell *et al.* 2007). In order to achieve high-quality results, each item or question should:

- contain clear and simple language;
- avoid superfluous information;
- avoid hints or clues directed towards the correct answer;
- contain suitable distractors (incorrect answers) to ensure the question is challenging;
- be relevant with regards to the aim of the questionnaire.

In addition to these points, questions were reviewed in consultation with peers and professors to ensure that they covered relevant topics, and that non-trivial information could be gained from student responses. By following these guidelines we can be confident that the questions test what we expect them to, and that through consultation with experts the test as a whole gains a measure of validity.

Statistical tests and considerations

Reliability, consistency and links between different conceptual groups

Links between different areas of conceptual knowledge and self-evaluation can be tested using

the Pearson biserial coefficient, which measures how well-correlated the performance is in different areas. However, this is a weak measurement for this questionnaire because answers are mostly true/false, and a large amount of noise can be caused by students guessing the correct answer. Cronbach's alpha is also used to measure how internally consistent each question group is. The QMDQ tests several different areas, which may not be linked, and within each group questions of a different style are asked. Hence low scores for correlation and consistency do not necessarily imply poor question design.

In addition to ensuring that this questionnaire has an acceptable validity, the reliability of the questionnaire will also be tested. Good reliability indicates that the questionnaire is able to distinguish students of different abilities. Using item difficulty and item discrimination index, individual questions can be reviewed and Ferguson's delta can give a measure of the questionnaire's over-all reliability. Reliability tests can only be used on the conceptual part of the survey as the first part gives no score to students' answers.

Bias

When sampling, bias is introduced and the results of a survey may not be representative of an entire population (Anderson *et al.* 2009). Completion of the QMDQ will be voluntary for students, which can potentially introduce bias because sampling is not random and the entire class is unlikely to participate. Volunteer sampling is used because it is a good method of obtaining sufficient numbers of willing participants. However, because the sample is no longer random, the results may be unrepresentative of the population as a whole.

Response bias is introduced from the way in which students respond to a question. For example, the way in which a question is worded can have an effect on the respondent's answer. When writing questions the principles outlined above will be followed to ensure that the language used is neutral and does not influence the decisions students make.

Non-response bias is caused by a fraction of the population not responding to the questionnaire. This applies to this investigation because we are using volunteer sampling and do not expect all students to respond. We are unable to judge how strong this bias will be or in what direction it will skew our results. Non-response bias could skew towards higher scoring students if they are more willing to participate or, conversely, high-scoring students may have a heavier workload and be less likely to participate. Without a parallel random

sampling, which was not feasible for this preliminary project, it is not possible to know how this bias will affect our results. Students will be encouraged to participate through announcements in lectures and email reminders, as a large number of participants will reduce the effect of non-response bias and improve the data for statistical analysis.

Questionnaire

One aim of the questionnaire is to identify gaps or weaknesses in student understanding of time-dependence in quantum mechanics and other related topics. Topics were chosen from the learning outcomes of the two second year courses discussed above, to ensure all students had covered the areas being tested. This prevents any bias being introduced from potentially different experience at an undergraduate level.

The self-evaluation of confidence and learning style consists of 10 questions that were designed for clarity and focus. Students were asked to indicate whether they understood quantum mechanics via a mathematical or visual approach. Response choices for these questions were formatted on a four-point Likert scale such that no neutral response was available. This method was chosen to force students to state a preference.

The QMDQ was written using LimeSurvey with embedded LaTeX equations. Students were informed about the questionnaire, and why we hoped it could be useful, in a lecture and via email. Reminders were sent a few times. Whilst the reasons for the questionnaire were communicated to the students, it remained voluntary and full participation was not expected. Students remained anonymous throughout the analysis. This is important as students were informed that their participation and performance on the survey would have no effect on any other aspects of their degree.

Students' responses were collected and analysed using the R programming language. We collected the results and found 47 of the 137 students (34%) had participated. Clearly, this is only a section of the entire population and hence any conclusion drawn from results will be indicative, but not fully representative, of the entire group.

For the analysis of the survey, Questions 11–30, the conceptual part of the questionnaire, have been split into items 1–33 to enable marking and analysis in R. The number of items has increased due to several two-part true/false questions which have been assessed individually. (See below for a discussion how this could be improved.) The self-evaluation question section is assessed

separately and has the same Q1-Q10 notation as occurs on the QMDQ.

Below is a set of typical sample questions from the questionnaire:

- Qn 1: I am confident in my understanding of the underlying concepts of quantum mechanics, e.g., wave particle duality and the uncertainty principle. (Agree/Slightly agree/Slightly disagree/Disagree)
- Qn 7: When answering QM questions, which area gave you most difficulty? (Remembering mathematical formulae/Visualising a suitable representation of the quantum mechanical system/Knowing how to apply the mathematical formulae/none of the above)
- Qn 13: In a double slit experiment, light shines through two narrow slits to create an interference pattern on a screen. The light is replaced by an electron beam, and electrons are fired through the slits one at a time. What is observed on the screen? (Two fringes are observed corresponding to the two slits/An interference pattern similar to that for light is observed after the first electron hits the screen/An interference pattern similar to that for light is observed after many electrons have hit the screen/The electrons arrive at random positions, with no discernible pattern.)
- Qn 20: If $\hat{A}\psi(x) = a\psi(x)^2$, consider the following two statements: i) $\psi(x)$ is an eigenvector of \hat{A} ; ii) a is an eigenvalue of the operator \hat{A} . [4 true/false choices]

Reliability of the QMDQ

The first set of analysis on the results from the QMDQ will study the reliability of the questionnaire. It will examine the conceptual part of the survey as the self-evaluation part has no correct or incorrect answers. Determination of good reliability proves that the questionnaire is able to discriminate between students (Lord 1952). The three tests of reliability used here are item difficulty index, item discrimination index and Ferguson's delta. The first two statistics test the reliability of individual questions and Ferguson's delta tests the reliability of the survey as a whole (Cohen 1988).

Item difficulty index

The item difficulty index is simply a measure of how students performed on each item, the average student score. A question is considered to be too easy if the item difficulty index has a value greater than 0.9 and too difficult for values of 0.3 or less.

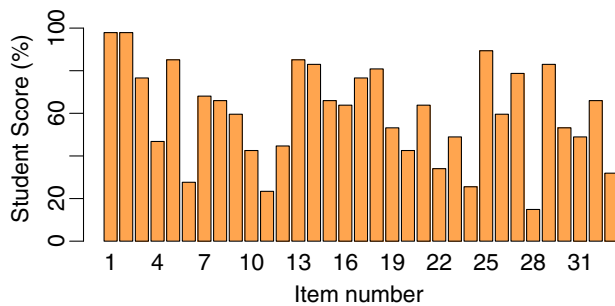


Figure 1 Average Student Score (%) for each question, items 1–33

Figure 1 shows the item difficulty index (average student score) for each question. By this measure, items 6, 11, 24 and 28 are too difficult and items 1 and 2 are too easy. The mean item difficulty on each item was 0.60 with a standard deviation of 0.10. This is an appropriate result; the test is difficult but not to the extent that student performance cannot be judged.

Item discrimination index

The item discrimination index is a measure of how good a question is at distinguishing between a high scoring and a low scoring student. We first divide the students into quartiles based on their overall score; the discrimination index for a question is then calculated as the percentage of the upper quartile students who answered the item correctly, minus the percentage of the lower quartile students who also answered the question correctly.

It has a possible range of -100 to $+100$, with larger values indicating a better discrimination. An item discrimination >30 is supposed to indicate a useful question (Adams & Wieman 2011). The discrimination index of each question is shown in Figure 2. The mean value of the item discrimination index for all of the questions is 24.6 (18.8). Overall

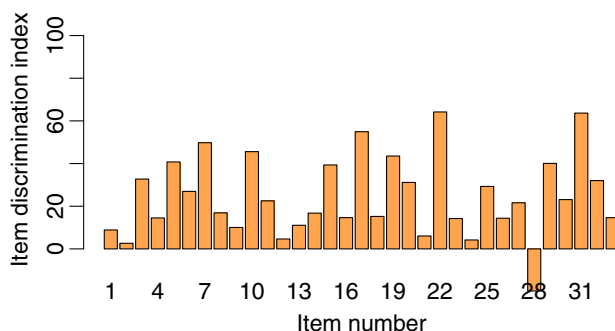


Figure 2 Item discrimination index for each question, items 1–33. The item discrimination index measures the items ability to distinguish between upper and lower quartile students. A value of 30 or greater is considered an acceptable score for the item discrimination index

the conceptual part of the questionnaire was reliable in its discriminatory abilities. A few questions did not perform well having an item discrimination index of less than 15 each. These questions have been reviewed to check their suitability and the possible reasons behind their poor performance.

Ferguson's delta

Ferguson's delta is an evaluation of the discriminating measure of the test as a whole, as indicated by the ratio of different test results to the greatest number of different results the test could generate. A test is considered to discriminate usefully for a value of 0.9 or greater.

The Ferguson's delta for the conceptual part of the test was 0.92 (0.42). One of the aims of this conceptual questionnaire is to distinguish between possible groups of students and this value indicates that the first draft of the QMDQ does this well.

Analysis

In order to test the relationship between different conceptual areas a correlation analysis was performed which calculated the Pearson correlation coefficient. This coefficient between two variables is the co-variance of the two variables divided by the product of their standard deviation. The definitions of suitable values for the Pearson coefficient vary depending upon the type of experiment being performed and the level of noise that can be expected in the results. Values for this correlation are much lower in surveys, for example where a strong correlation would be indicated by a value of 0.5 or higher (Cohen 1988). Figure 3 shows that time-dependent questions have a poor correlation with the other areas of the survey: probability, eigenfunction and concepts in quantum mechanics.

The internal consistency of the survey was judged using Cronbach's alpha (Cortina 1993) with the overall consistency measuring 0.436, less than the 0.5 required for an acceptable value. Consistent performance across all items indicates a test is reliable in discriminating between candidates.

However, as the QMDQ was designed with the aim of testing students' misunderstandings, this may have contributed to the low consistency. The internal consistency of these individual groups was also low. The overall consistency may be improved in the review of the QMDQ by tightening question bearing and removing questions that do not coincide with the different groups or overall theme of the questionnaire. Purposeful difficulty changes in groups and across the questionnaire, in order to

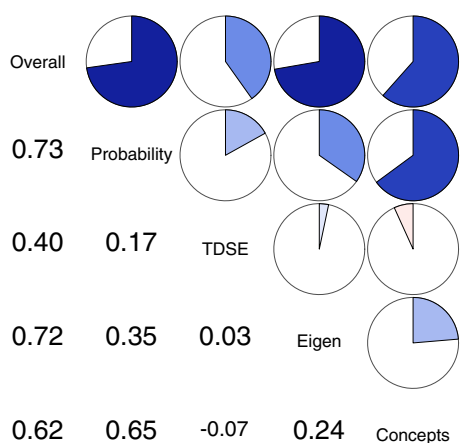


Figure 3 A corelogram, a matrix of Pearson correlation coefficient results, between groups and linked to overall performance. The numerical value and corresponding pie chart is shown for each correlation

test student understanding, may be the cause of the poor internal consistency of the questionnaire.

The first ten questions of the QMDQ surveyed student confidence as well as students' preference between understanding quantum mechanics with mathematics or intuitive content. Both were measured with a four-point Likert scale with positive or negative weight representing each choice. One possible explanation for poor student performance in time-dependent quantum mechanics questions was low confidence, causing a reluctance to attempt seemingly more complex questions. Confidence was tested with Q1, Q2, Q6, Q9 and Q10. With the Likert scale students who declared to be strongly or slightly confident are marked with a score of +2 or +1 respectively. Conversely, students who asserted that they were not confident are correspondingly awarded -2 or -1. This gives a possible range of responses from -10 to 10 or very low confidence to very high confidence. The mean confidence score for all students was 5.2 (3.4). Figure 4 shows that confidence is positively correlated with student performance on the conceptual part of the test with

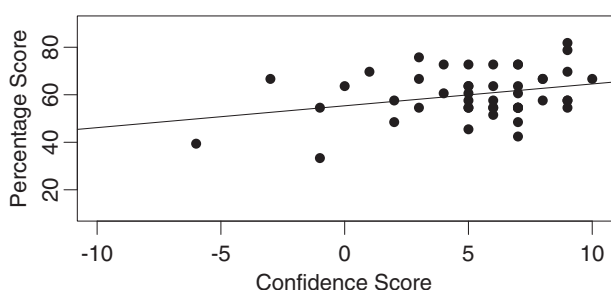


Figure 4 Student performance (%) against confidence. The linear regression is also shown

a gradient of 0.9, standard deviation of 0.4 and a p-value of 0.036.

The effects of students' preferences for either a mathematical or a descriptive approach to quantum mechanics were also investigated in order to study the effect of a learning style on subsequent performance on the same test. No correlation is observed between student score and student preference of learning style. Teaching methods that vary between mathematical and descriptive understanding have been reviewed in other studies and some favour descriptive methods to build physical intuition about quantum phenomena (Meyer and Land 2005).

In addition to the statistical measures outlined above, a p-value measures the probability that a result is based on random choices (Wright 1992). Conventionally a p-value of less than 0.05 is acceptable as this gives a better than 95% certainty of a non-random result. Due to the use of true/false questions having a mean score of 57%, many of the items on the survey have poor p-values. This does not indicate students were guessing but we cannot statistically say that they were not random guessing in many items. Nor do the poor p-values indicate a problem with the question content, but they do highlight the disadvantage of using true/false questions. Several true/false questions had mean scores of less than 30% while others had mean scores greater than 80% resulting in p-values of less than 0.05. Despite the poor p-values on some questions, we can see that students were not guessing on other questions, hence we do not expect random selection to have a major impact overall.

Final version of the QMDQ

The revised version of the QMDQ is a diagnostic tool capable of measuring students' confidence in quantum mechanics, preferential learning style and knowledge of quantum mechanics. The revisions to the conceptual part of the test are outlined above. Changes were implemented in order to learn more about student understanding and refine the quality and reliability of the questionnaire. New questions were introduced and some items were removed with the aim of producing a questionnaire that is reliable, discriminatory and concise. The self-evaluation section of the questionnaire was also reviewed and students were asked to evaluate their confidence and preferences in relation to their current course PHYS 30101. These alterations allow the comparison of student confidence and learning style preference for different courses, in addition to

the changes in student performance on the conceptual part of the survey.

New 4-part multiple choice questions were introduced into the conceptual part of the survey in order to gather more information on areas in which students performed poorly. Several other questions were re-arranged so that related pairs of true/false items are contained within the same question. This restructuring allows a change in the marking. In order to obtain results with less statistical noise, a student must answer both (closely related) true/false questions correctly to obtain a full score; a reduced score will be given for an incorrect answer. This may lead to improved p-values for these questions compared with singular true/false items. The difficulty of these combined item questions is hard to judge and may require some further fine-tuning with a sample group. The original and new multiple choice questions are expected to remain at the same level of difficulty and to show an improved discrimination between students. The remaining true/false items are included to measure any change in student performance and highlight areas of particular fundamental knowledge. The true/false questions will, however, retain poor p-values due to the statistical noise of a two-option question.

This revision of the QMDQ can be analysed in much the same way as the original. Sections that have been repeated can be compared and areas of improvement and stagnation in student scores highlighted. The reliability, grouping and p-values can also be compared to confirm improvements in the questionnaire or weakness in student understanding. The performance of the few new questions and some significant changes will also be examined critically. The KR-20 test could also be used to compare the two data sets and determine their similarity. Evaluation of question groups using the Pearson coefficient may show an improvement as a consequence of reduced statistical noise in the new question structure in the QMDQ.

It was not possible to obtain results for the revised QMDQ within the timeframe allowed for this project. The revised survey was written and submitted to students but an inadequate number of responses was returned for analysis. In future studies, the revised version of the QMDQ can be used to examine student performance and knowledge level over the subjects covered, as a basis for a more detailed questionnaire or learning resource.

Conclusions

This project was completed successfully with the preparation and implementation of a quantum

mechanics diagnostic questionnaire (QMDQ). From the results of the initial version of the QMDQ we have been able to analyse the reliability of each item in the conceptual part of the questionnaire. The mean score was 60% for the 47 students who participated and the Ferguson's delta for the test as a whole was 0.92 (0.42), which indicates that the test can distinguish between students well. The mean item discrimination index was found to be 24.6 (18.8), less than the desired value of 30. Revisions to the survey have been made in an effort to improve the reliability of the questionnaire and the strength of the conclusions that can be drawn from it. Poor questions were removed or adjusted and new questions were written to further explore interesting aspects of student ability. More multiple-choice questions were introduced and true/false questions were linked more closely with the aim of improving the grouping of questions and to permit more detailed conclusions to be drawn from student answers.

A positive correlation between student confidence and the results from the conceptual questions has been found. There was no correlation detected between students' preferred teaching style and the results from the conceptual questions. Time-dependence was found to be weakly correlated to student performance in other topics in quantum mechanics. This provides some confirmation that some students have a 'fear' of time-dependence which prevents them from reaching their expected performance. All of the questions used in the QMDQ were analysed, and several important improvements were made to create a revised, but not yet tested, QMDQ.

Students were found to perform poorly on items 19, 23 and 28. These questions involved the use of wavefunctions and tested basic knowledge of probability and time-dependence for different examples. In order to probe this potential misconception further two new multiple-choice questions are included, covering both probability density and time-dependence. This should allow us to say precisely how many students struggle with the concepts of how different solutions to the TDSE affect subsequent time-dependence and certainty of different results.

The number of correct responses to item 4 was lower than expected; despite a poor discrimination, this question is preserved. The question is clear and of a high quality. Therefore poor performance here suggests that students across all ability levels struggle with the wave-particle duality and uncertainty in position of quantum objects. Though unrelated to the observed weakness in time-dependence this performance does represent a general hesitance for students when they move

away from the classical areas of under-graduate physics. Other questions in the QM concepts group were answered well, with the exception of a misconception that a quantum object loses energy when tunnelling.

Time-dependent questions scored poorly, with a mean score of 47% compared to 60% for the entire survey. In items 11, 24 and 28 students scored less than 25%. These questions examine student understanding of solutions to the TDSE, separation of variables and linear superposition of wavefunctions. The time-dependent properties and process of separation of variables are covered in detail in PHYS 20101. Students may have struggled to grasp the concepts behind the straightforward mathematics, leading to a poor retention of knowledge.

Areas of time-dependence that students were able to answer well included remembering that the solution of the TDSE produced eigenvalue equations for the time and space dependent parts of the wavefunction. In order to address the weaknesses highlighted, two new multiple-choice questions are included in the final version of the QMDQ which cover understanding of different wavefunctions.

Discussion

Through discussions and meetings with academics in the field of quantum physics, a diagnostic questionnaire was created and tested on students. The QMDQ was then refined based on what we were able to learn from the results of the first set of students. New questions were written to study the areas in which students struggled and some questions were removed if they were unable to add to the survey, either due to repetition or irrelevance. The updated QMDQ is a tool that can be used to judge students' ability in time-dependent quantum mechanics and other areas. The questionnaire will also judge confidence and preferred learning style. The QMDQ is able to give a quick assessment of ability and a rough indication of where students begin to lose confidence in answering time-dependent questions.

The first version of the questionnaire was useful and highlighted several areas of interest. There were, however, several ways in which we learned that it could be improved. Through analysis it became clear that the true/false questions resulted in some statistical noise, which resulted in poor p-values and low consistency. This issue has been addressed through the implementation of more multiple-choice questions and linked true/false questions. For the items to be linked they must

both test the same idea, but without asking the same question. Linked items can then be marked such that for a student to gain full marks, they must answer both parts correctly, thereby showing understanding of the concept being tested.

This project has resulted in the creation of a diagnostic tool that has been developed so as to have an improved validity and reliability. This tool has the potential to be used in future projects as a measurement of student understanding at different stages of their studies. Any such future project could also use interviews to study student understanding and act as a learning resource. There are several methods for student interviews that can be considered depending on time restrictions and objectives.

Future direction

Even at this level, the quantum mechanical concepts of tunnelling and uncollapsed wavefunctions were found to cause difficulty and are therefore good topics to consider for future study. These quantum mechanical concept questions could be used as the introduction or baseline for a student interview before moving on to the more complex time-dependent problems. Alternately, a new conceptual questionnaire could be built, targeting the areas in which the QMDQ found poor student performance.

Time-dependent questions revealed that students remembered learning the process of separating variables but did not retain much of the knowledge as to what the solutions represent, or how different solutions display different time-dependent properties. In order to locate the point where students lose confidence, a structured interview could be designed where the TDSE is solved and subsequent solutions are manipulated. Through this process it would be possible to pinpoint where each student begins to struggle, either with the mathematics or the conceptual understanding.

The structured interview could also be used to help students expand their knowledge and interviewers could help students overcome areas of uncertainty or misconception. The success of the interviews could be measured by the improvement in students' scores on the QMDQ and the findings that locate the conceptual barriers to student progress. If the improvement of student understanding is of a higher priority, group discussions or problem sheets could be designed with a focus on helping students understand the concepts behind the questions which received low scores.

Acknowledgements

We thank Professor Mike Birse for his critical support during this work. This report was written

by N.R. Walet based on the research done and reports written by J.R. Chapman and S.O. Hunt as part of a 4th year project under his supervision.

References

- Adams, W.K. and Wieman, C.E. (2011) Development and validation of instruments to measure learning of expert like thinking. *International Journal of Science Education* **33** (9), 1289–1312
- Anderson, R.S., et al. (2009). *Statistical methods for comparative studies: techniques for bias reduction*. Vol. **170**. Wiley-Interscience.
- Aubrecht, J.G. and Aubrecht, J.D. (1983) Constructing objective tests. *American Journal of Physics* **51** (7), 613–620.
- Cataloglu, E. and Robinett, R.W. (2002) Testing the development of student conceptual and visualization understanding in quantum mechanics through the undergraduate career. *American Journal of Physics* **70**, 238.
- Cortina, J.M. (1993) What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology* **78** (1), 98.
- Cohen, J. (1988) *Statistical power analysis for the behavioral sciences* (2nd ed.) Routledge.
- Cramer, J.G. (1986) The transactional interpretation of quantum mechanics. *Reviews of Modern Physics* **58**, 647.
- Edinburgh archive of conceptual tests in physics. Available at <http://bit.ly/ConceptTests> (accessed 14 July 2013).
- Falk, J. (2007) *Students' depictions of quantum mechanics: a contemporary review and some implications for research and teaching*. (PhD dissertation) Uppsala University.
- Feynman, R.P. (1990a) *The Nature of Physical Law*. London: Penguin.
- Feynman, R.P. (1990b) *QED: The Strange Theory of Light and Matter*. London: Penguin.
- Lindell, R.S., Peak, E. and Foster, T.M. (2007) Are They All Created Equal? A Comparison of Different Concept Inventory Development Methodologies. In *AIP Conference Proceedings* **883**, pp14–17 (eds. L. McCullough, P. Heron and L. Hsu). Melville, NY: AIP Press.
- Lord, F.M. (1952) The relationship of the reliability of multiple-choice test to the distribution of item difficulties, *Psychometrika* **18**, 181–194.
- Maloney, D.P., O'Kuma, T.L., Hieggelke, C.J. and Heuvelen, A. (2001) Surveying students' conceptual knowledge of electricity and magnetism. *American Journal of Physics* **69**, S12–23.
- Meyer, J. and Land, R. (2005) Threshold concepts and troublesome knowledge (2): Epistemological considerations and a conceptual framework for teaching and learning. *Higher Education* **49** (3), 373–388.
- Sharma, M. (2013) Private communication.
- Singh, C. (2001) Student understanding of quantum mechanics. *American Journal of Physics* **69**, 885.
- Singh, C., Belloni, M. and Christian, W. (2006) Improving students' understanding of quantum mechanics. *Physics Today* **59** (8), 43–49.
- Styer, D.F. (1996) Common misconceptions regarding quantum mechanics. *American Journal of Physics* **64**, c31.
- Wright, S.P. (1993) Adjusted p-values for simultaneous inference. *Biometrics*, 1005–1013.
- Wuttiprom, S., Sharma, M.D., Johnston, I.D., Chitaree, R. and Soankwan, C. (2009) Development and use of a conceptual survey in introductory quantum physics. *International Journal of Science Education* **31** (5), 631–654.