Evaluation of Student Understanding of Uncertainty in Level 1 Undergraduate Physics Laboratories

Philippa L Petts[1], G. Peter Swift[2], Samuel C Nightingale[3]

Department of Physics, Durham University, United Kingdom[1]
Department of Physics, Durham University, United Kingdom[2]
Department of Physics, Durham University, United Kingdom[3]

Abstract

In recent years, work has been undertaken at Durham University to investigate how student understanding of uncertainty in experimental measurements changed throughout the course of the instruction they had during their first year of undergraduate laboratories. Measurement uncertainty can be thought of as a threshold concept, which students need to pass through before they gain a full understanding of both measurements and uncertainties, allowing them to transition from the point to the set paradigm. Students were surveyed both pre and post instruction to see how their understanding had developed and deepened throughout the course of the year of study. A particular focus was placed on assessing student understanding of repeat measurements, and the agreement of measurements. Following the period of instruction, the sophistication of answers provided by students to the survey, showed a statistically significant improvement as measured by an optimal pooled t-test, as the survey responses consisted of a mixture of both unpaired and paired data. However, misconceptions still remained, particularly when looking at students' understanding of repeat measurements.

Keywords: *Physics Education; Undergraduate Experimentation; Uncertainties*

1. Introduction

Laboratory work has long been held as playing a central role in scientific study. Griffin purported in 1892 that '*laboratory has won its place in school; its introduction has proved successful. It is designed to revolutionise education*' [1]. Science education was then heavily influenced by a progressive movement, promoting a learning by doing approach. In the 1960s, laboratories became re-established as a core part of the scientific process. They have a number of goals including: to arouse interest and curiosity in science; the development of scientific thinking and the scientific method; and to develop practical abilities [2]. In addition to the above, laboratories in the early stages of a degree also focus on concepts and models, but are often less concerned with a procedural understanding. Millar *et al.* break down procedural understanding into three areas – the purpose of performing an experiment, the ability to manipulate laboratory equipment and having an appreciation for the reliability of a set of results [3]. In the twenty-first century, we hope that our teaching enables students to adopt a constructive approach to learning, allowing the learner to accommodate and assimilate new knowledge.

1.1 Threshold Concepts

Threshold concepts can be thought of as 'conceptual gateways:' an academic hurdle that, once cleared, opens up a previously inaccessible way of thinking. Meyer reports that threshold concepts are distinguished by five key criteria [4]:

- Transformative – once understood, there is a significant shift in subject perception;
- Integrative – a threshold concept exposes previously hidden interrelatedness;
- Irreversible – it is unlikely to be forgotten without significant effort;
- Boundary-defining – it will likely outline a specific conceptual space and serve a limited purpose;
- Potentially troublesome to learn – the knowledge is usually counter-intuitive.

If a student has an incomplete understanding of a threshold concept it is likely to have long-lasting repercussions, forming a barrier, preventing application in any unfamiliar contexts and stunting their further educational progress.

Measurement uncertainty is central to experimental Physics; students having a thorough understanding being widely regarded as essential for strong academic progress in this field [5]. The ability for students to assess the reliability of a set of measurements and use this information to guide experimental procedure is a key skill for achieving success in the laboratory. Lubben, Buffer, Allie and Campbell characterise data handling via the point and set paradigms, defined below [6]:

The point paradigm is characterised by the underlying notion that each measurement could in principle be the true value. As a consequence each measurement is independent of the others and the individual measurements are not combined in any way.
The set paradigm is characterised by the notion that each measurement is only an approximation to the true value and that the deviation from the true value is random. As a consequence, a number of measurements are required to form a distribution that clusters around some particular value.

Students can be between these two paradigms, in what is known as a state of liminality as described by Meyer and Land. The work of Lubben *et al.* reports that, whilst students generally display characteristics of both paradigms, the overall goal of teaching should be to move students wholly into the set paradigm, and in particular to be able to give their reasoning in the set paradigm. For example, they would be able to reason, in an appropriate way, why the mean is the best estimate of a quantity's true value.

Meyer's criteria for threshold concepts can be applied to show that measurement uncertainty is one such concept; students need to pass the threshold to fully reason and understand this topic. Below explains how measurement uncertainty fits into the definition of such a concept:

- Transformative – A good understanding of measurement uncertainty will transform a student's thought process and success in a laboratory setting. When the threshold is crossed, uncertainties become seen as intrinsic to measurements, and are an assessment on data quality;
- Integrative – many previously studied concepts must be combined and fully understood when looking at measurement uncertainty, for example random and systematic error, repeat measurements;
- Irreversible – once fully grasped, students will see data as a spread of results, with the uncertainty characterising the data's reliability;
- Boundary-defining – a more general approach must be taken, regarding the topic as a boundary to quantitative analysis;
- Troublesome – accepting that measurements are inherently uncertain, and that there is no perfect scientific process troubles physicists, who start their academic journey from the position of requiring exact statements, rules and clarity.

In this work, an investigation into measurement uncertainty as a threshold concept is reported, in the context of Level 1 (first-year honours) undergraduate physicists at Durham University. The definitions of Hughes and Hase (2010) [7] are used: *'an accurate measurement is one in which the results of the experiment are in agreement with the 'accepted' value.'* whilst a *'precise measurement is one where the spread of results is 'small', either relative to the average result or in absolute magnitude'*. Hence accuracy is affected by systematic error and precision by random error, and in experimental work, random errors should be minimised.

2. Methodology

Eight probes, surveying a variety of measurement uncertainties, were given to undergraduate students at Durham University registered in the Discovery Skills module, PHYS101 during the 2019-20 academic year. This laboratory-based module is taken by all first year physics students and also students studying for a degree in Natural Sciences from both the first and second year. It also includes a lecture-based course on error analysis, with learning of this topic supported through both problem sets and as specific tasks completed within the practical laboratories. The survey was given before any instruction had been given, when students' only prior knowledge of experimental work and measurement uncertainty will be from UK A Levels, or equivalent. The probes chosen were based largely on those of Lubben and Miller (1994) and aimed to assess student understanding of a variety of key issues that arise in measurement and uncertainty work [8]. As a number of the probes were related, with information in some of the later probes potentially having the effect of making it easier to answer some of the early probes more fully, the survey was given in such a way that the questions had to be answered in a fixed order, and once an answer was completed, it could not be altered. Consequently, the results should more likely reflect the deep understanding of students, rather than what they are able to repeat when prompted.

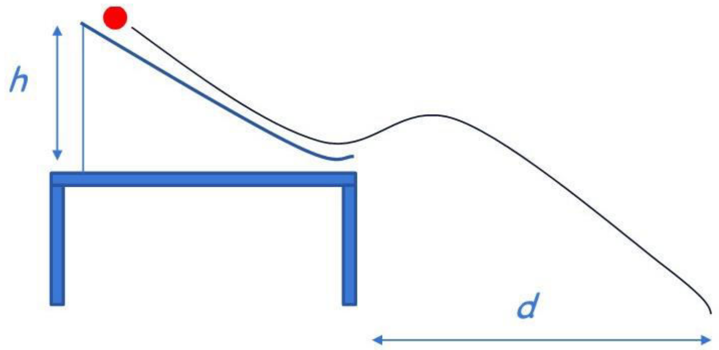

Figure 1: The set-up that was used in the survey's experiments, with the ball dropped from a height h, before travelling a horizontal distance d, which is then measured.

The probes all considered an experiment where groups of students roll a ball down a slope, clamped to the edge of a table. The ball is released from a height h, leaves the table horizontally, and then falls on the floor some distance, d away as shown in Figure 1. The probes test a variety of measurement and uncertainty phenomena, at a level appropriate for first-year physics students. For example, looking at reasons for making repeat measurements, how final results should be presented, and then the notion of agreement between values obtained from the same experiment by two independent groups. All questions, provided a multiple-choice answer that students selected, followed by a free-text area for reasoning to be explained.

Following completion of the survey by students prior to their teaching instruction, responses were coded thematically using a modified version of a scheme developed over previous years. For student responses covering key themes in multiple categories, a 'tick all that apply' approach was used in order to best reflect the overall trend of student thought. Following completion of the pre-test, four students from the year group took part in thirty minute interviews, to enable the coder to more fully understand student responses and to allow a more detailed probing into key misunderstandings. In addition, during the interviews students' prior knowledge was also discussed from their learning during, for example, A-Levels at school. In particular, this allowed a probing of their knowledge of the key terms already mentioned, namely 'accuracy' and 'precision'. Towards the end of the academic year when the students had undertaken all of the work in support of the learning outcomes associated with measurement and uncertainty, students were asked to complete the same survey again. The responses to the post-survey were then thematically analysed in the same way as those for the pre-instruction survey.

To compare the data between pre and post-tests, seven of the answers were given a numeric score in order to investigate overall learning gain. The scores were based on the explanation's sophistication: a mark of one was awarded, if the answer surpassed a given sophistication level, and zero was given otherwise. For example, when looking at the agreement of two sets of results, low sophistication would correspond to "their means are close/far apart," whilst a high level of sophistication would be represented by answers like "the measurement means are not in agreement as they don't lie within three standard errors of each other." This allowed the mean scores and associated standard deviations to be found for the whole survey and for specific sections of it. Because this data was a mix of paired and unpaired data (some students had completed both surveys and others only either the pre or post-survey), scores were compared using an optimal pooled t-test [9], providing a weighting inversely proportional to the variances of the estimates. The mean scores of the pre and post-test, μ_1 and μ_2 were calculated, and the weighted difference in these means was then compared, through the test statistic, $T_0 = \mu/S_\mu$ to the relevant t-distribution table.

3. Results and Discussion

A total of 90 students completed the pre-survey, 51 the post-survey whilst 20 completed both, with a similar gender split being seen across both surveys. Around 80 % of the students were registered on the Single Honours Physics programme, with the remaining 20 % comprised of the students undertaking this module whilst reading for a degree in Natural Sciences.

3.1 Repeat Measurement Probe

The first two probes focussed on the making of repeat measurements and whether students would choose to do this or not. In these probes, students were asked to decide which statement they agreed with from: taking a few more measurements; taking one more measurement; or taking no more measurements, and then to explain their answer. In the pre-instruction survey, irrespective of which statement the students chose, six main ideas were provided for the reasoning as shown in table 1. Following instruction student responses were seen to improve, with 33 % of students being able to link the idea of repeating measurements to obtaining the standard error, whilst no student mentioned this in the pre-test. The general sophistication of written answers improved from statements like ‘so that a mean can be calculated’ to ‘the more measurements taken, the smaller the standard error.’ Interestingly, in this probe a large number of students still argued that repeats are needed to improve accuracy!

Category	Description	Pre-Test	Post-Test	Change (%)
RD 1	Repeats needed to calculate mean	50	43	-7
RD 2	Repeats needed to improve accuracy	29	18	-11
RD 3	Repeats needed to identify anomalies	31	16	-15
RD 4	Repeats needed to reduced uncertainty	41	28	-13
RD 5	Repeats needed to find/increase standard error	0	33	+33

Table 1: Summary of responses to the Repeat Distance (RD) probe, including the level of change seen between pre and post instruction testing.

3.2 Probes on Agreement of Data

Four probes were considered here. Students were presented with two sets of data and asked to decide whether the measurements agreed or not, perhaps including having to complete a calculation. Within the probes, some had the same mean, whilst others had different means, but these datasets had similar or different spreads. These probes got more complex as they progressed, and even in the pre-instruction test, student responses became more sophisticated in later questions, since each successive question gave more information, for example the term standard deviation is introduced for the first time in the final probe, and it is at this point a student potentially realises that they have made a mistake in the earlier probes. However, when comparing written answers given in the post-instruction survey, student answers on each probe where in general seen to be more sophisticated than in the pre-test survey. Table 2 shows the improvement in responses for both the ‘Different Mean Same Spread’ (DMSS) and ‘Different Mean Outside Spread’ (DMOS) probes. Here, outside spread refers to the fact that the mean of neither data set lies within the spread of the other data set.

Category	Description	DMSS Change	DMOS Change
DM 1	It depends on how close the means are numerically	-46	-28
DM 2	It depends on the means and uncertainty	+21	+4
DM 3	It depends on the means and standard deviations/errors	+31	+38

Table 2: Percentage change for the DMSS and DMOS probes between the pre- and post-surveys.

It is from these two probes in the pre-survey that it is possible to draw the most interesting conclusions regarding student misunderstanding of uncertainties. With over half the students on both probes offering simple numerical reasoning, it is clear that students commencing their undergraduate studies have very little appreciation of the content they are soon to learn. By the time of the post survey, all of these probes showed a significant improvement in the sophistication level of responses.

3.3 Statistical Learning Gain

The average of the raw scores obtained by students on the overall test, and also on the questions specifically looking at repeat measurements and levels of agreement, are shown in table 3. The T_o value is also given, with the level of significance which this represents, using the optimal pooled t-test. This shows that, to a significance level of less than 0.1%, students’ scores on the survey improved, but that the majority of this score increase comes from increases in the sophistication of student answers to the probes on agreement level rather than on those considering repetition of measurement. Indeed, the small improvement in the scores for the repeat measurement probes was not even significant at the 10% confidence level. From this, it can be concluded that the answers given by students to the post-test questions on agreement are much more sophisticated, shifting towards standard deviation and error away from a purely numeric comparison. However, despite this large improvement shown in the scores, only around half of the students provided what could be termed a model answer, showing that they had fully moved through the threshold. Correspondingly, for the probes on repeat measurements, many students are still in the liminal state after nearly a year of teaching.

	Pre-Test	Post-Test	T_0	P
Total (7)	2.5 ± 1.4	4.2 ± 1.5	6.5	< 0.001
Repeat (2)	0.6 ± 0.7	0.9 ± 0.8	1.1	> 0.1
Agreement (4)	1.8 ± 1.0	3.1 ± 1.0	6.0	<0.001

Table 3: The raw scores and their standard deviation for the pre-survey and the post-survey, split into overall score and scores on the repeat measurements and measurement agreement sections. The T_0 value from the optimal pooled t-test is included alongside the associated P value.

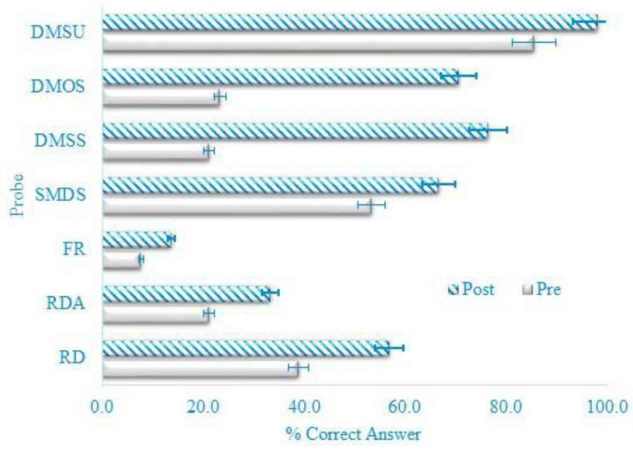

Figure 2: Comparison of the percentage of students to give correct answers on each prove in each survey. Error bars are given at the 95% confidence interval.

By only considering only the paired sets of data, and using a binomial exact test, a similar trend to the increase of the sophistication of the responses given can be seen in figure 2. In particular, it highlights an improvement in percentage of correct answers on all probes, specifically the DMSS and DMOS probes.

5. Conclusions

First year students at Durham University had their understanding of measurement and uncertainty tested both prior to and following instruction. This topic was chosen because it has been postulated to be a *Threshold Concept* for physics students. After the period of instruction, there was a statistically significant improvement in the performance of students, in particular with regards to the agreement of two similar datasets. This suggests that overall, the learning undertaken in the first year undergraduate module is somewhat successful in transitioning students from the 'point paradigm' to the 'set paradigm' associated with threshold concepts. However, it is clear that many students have not reached a satisfactory level of understanding or exist in a state of liminality. Consequently, further work will be undertaken to enhance the teaching of these key topics at Durham in support of deeper student learning.

References

[1] Rosen, S., *A History of the physics laboratory in the American public schools (to 1910),* American Journal of Physics, 22, 194-204 (1954).

[2] Shulman, L. D. & Tamir, P., *Research on teaching in the natural sciences*, In R. M. W. Travers (Ed.), Second handbook of research on teaching, Rand McNally (1973).

[3] Millar, R., Gott, R., Lubben, F. & Duggan, S., *Children's performance of investigative tasks in science: a framework for considering progression,* M. Hughes (ed.), Progression in Learning (Clevedon: Multilingual Matters) (1996).

[4] Meyer, J., & Land, R., *Threshold Concepts and Troublesome Knowledge: Practising within the Disciplines*, (2003).

[5] Backerra, A. C. M., *Measurement Uncertainty as a Threshold Concept in Physics*, Physics Essays, *23*(3), 419–44 (2010).

[6] Lubben, F., Buffler, A., Allie, S., & Campbell, B, *Point and set reasoning in practical science measurement by entering university freshmen,* Science Education*,* 85, 311-327, (2001).

[7] Hughes and Hase, *Measurements and their Uncertainties*, Oxford University Press, (2010).

[8] Lubben, F. & Millar, R. *A Survey of the Understanding of Children aged 11-14 of Key Ideas about Evidence in Science* PACKS Research Paper 3, University of York, (1994).

[9] Guo, B., & Yuan, Y, A comparative review of methods for comparing means using partially paired data. *Statistical Methods in Medical Research*, 26(3), 1323–1340, (2017).