



PAPER • OPEN ACCESS

## Assessing first-year undergraduate physics students' laboratory practices: seeking to encourage research behaviours

To cite this article: K Dunnett *et al* 2019 *Eur. J. Phys.* **40** 015702

View the [article online](#) for updates and enhancements.



**IOP | ebooks**<sup>TM</sup>

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the [collection](#) - download the first chapter of every title for free.

# Assessing first-year undergraduate physics students' laboratory practices: seeking to encourage research behaviours

K Dunnett<sup>1,4</sup> , M N Gorman<sup>2,5</sup>  and P A Bartlett<sup>3</sup> 

<sup>1</sup> Nordita, KTH Royal Institute of Technology and Stockholm University, Roslagstullsbacken 23, SE-106 91 Stockholm, Sweden

<sup>2</sup> Department of Physics, Aberystwyth University, Penglais, Aberystwyth, Ceredigion, SY23 3BZ, United Kingdom

<sup>3</sup> Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, United Kingdom

E-mail: [paul.bartlett@ucl.ac.uk](mailto:paul.bartlett@ucl.ac.uk)

Received 19 June 2018, revised 2 November 2018

Accepted for publication 15 November 2018

Published 7 December 2018



CrossMark

## Abstract

Encouraging positive inquiry-focused behaviours within the constraints of a physics teaching laboratory environment can be challenging. Here, we report on an implementation, the 'working grade' (w-grade), designed to directly assess aspects of students' laboratory practice with the aim of encouraging first-year undergraduate students to look beyond the concept of a 'correct outcome' to a physics experiment. The w-grade is composed of the five aspects of group work, querying, exploration, attitude and progress which are each marked on a 0, 1, 2, 3 scale. The initial implementation is presented in full as well as a second, simpler variant. The w-grade emphasises and directly rewards inquiry behaviours and students were much more willing to explore the experiments than in previous years.

Supplementary material for this article is available [online](#)

<sup>4</sup> K Dunnett undertook this research study while at UCL prior to moving to the Nordic Institute for Theoretical Physics (Nordita).

<sup>5</sup> M N Gorman undertook this research study while at UCL prior to moving to Aberystwyth University.



Original content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](#). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

Keywords: education, practical physics, laboratory training, first-year undergraduate, soft skills

(Some figures may appear in colour only in the online journal)

## 1. Introduction

The purpose of laboratories in undergraduate science teaching has, and is, to a large extent influenced by pressures on resources (staff time, cost) as well as external pressures [1, 2]. Traditional ‘cookbook style’ laboratories have long been criticised for stifling cognition skills [3, 4], and it has recently been found to ‘high precision’ that (optional) laboratory classes that are designed to merely reinforce lecture content have ‘no added value’ [1, 2] which contradicts the assumption that laboratories can reinforce knowledge by allowing students to engage with particular content in different settings [3, 5].

Despite the criticisms, ‘cookbook’ style laboratories are the common form of laboratory instruction for pre-university students [6–8], including University College London (UCL) undergraduate physics students, the majority of whose prior experience of laboratory work is following a prescribed method [9]. It is often simpler for inexperienced demonstrators<sup>6</sup> to give the ‘correct’ answer to students’ questions rather than encouraging them to discover solutions independently. It has been found that in student-demonstrator interactions, the majority of discussions are concerned with low-level procedural issues specific to apparatus [10]. This can be due to inadequate demonstrator training, particularly for postgraduate researchers, who may use pedagogical practices that they experienced as students [11], and may have been discouraged from taking teaching duties seriously.

Inquiry-style laboratories are increasingly prevalent as part of research-based learning experiences in universities [12]. One innovation has been the design of ‘studio rooms’ in which traditional lectures and laboratories are combined into a shared space [13–17]. Another innovation, conducted at the University of British Columbia, is the structured quantitative inquiry lab [18], in which students are given relatively constrained experimental goal and set-up but they decide how to conduct the experiment and analyse data. From interviews with students it became apparent that the ‘sense of both agency and creativity contributed greatly to their enjoyment and motivation’. Another innovation is that at Rutgers University, the ‘Investigative Science Learning Environment’ [19, 20], in which students design their own experiments to investigate new phenomena, test hypotheses, make predictions and solve semi-realistic problems: in summary the ‘goal is to design an experiment whose outcome can be predicted based on the hypothesis to be tested’. Students are supported by heuristic guiding questions, self-assessment rubrics and reflection questions which are designed to provide ‘glass-box’ scaffolding [21] which supports students in internalising the implicit processes for any experiment as opposed to ‘black-box’ scaffolding in which they do not need to think.

One potential difficulty with inquiry-style laboratories (where students are free to modify an experiment to some extent [1]) is different student preferences or expectations that can affect their willingness to engage with open endedness [22–24]. In particular, students may prefer explicit instructions to complete well-defined tasks [22], an approach that allows constant comparison with peers [23]. Students can attribute failure to a lack of intrinsic ability [23] and may view questions and inquiry opportunities as a threat to self-esteem [24]. The desire to question demonstrators for the ‘right answer’ is believed to be linked to feelings of insecurity and a desire not to expose

<sup>6</sup> A laboratory demonstrator is a member of staff—academic, PhD student or other—present during the laboratory session whose rôle is to support students.

themselves [25]. Students may also think that seeking help can be interpreted as evidence for low ability [26]. In both cases, the students' behaviour may be driven by low self-esteem [25, 26] and there is a general need to encourage students to seek help and ask questions [27].

Educative assessment which requires students to actively engage with learning experiences that develop their understanding may be missing in many UK pre-university practical physics courses [28]. It appears that students think that assessment is predominantly to gain marks as opposed to being formative (on-going, habit-forming) as well as summative [29]. Also students can often lack clear ideas about the purposes of laboratory activities [30] and their perception of the purpose of laboratories does not necessarily align with the course designer's purpose [31, 32]. At UCL, new undergraduate students typically understand laboratories as 'teaching' or 'learning' skills or content—not as investigative environments [9].

At UCL, physics laboratories are not considered as a means to support theoretical studies but instead, experimental physics is treated as a discipline in its own right, and as a crucial element to the appreciation of physics as a 'way of approaching scientific discovery' [33]. UCL's practical training is designed to encourage students to fully explore experiments and the accompanying theory with discrepancies expected between the experimental results obtained by the students and theoretical predictions summarised in the scripts. We actively avoid the 'guided demonstration' approach to undergraduate laboratory physics.

Experiments are conducted over four 3.5 h laboratory sessions. All students attend the first term's weekly experimental sessions regardless of subsequent physics sub-specialisation (e.g. astrophysics). Students who follow an applied physics course then have two laboratory sessions per week in the second term and complete a further three experiments and a short electronics project. The laboratory scripts outline a potential experimental method, but students have to decide on the details of the procedure, which they are expected to iterate for repeat experiments, based on their own observations. Furthermore, the laboratory scripts contain comments and questions designed to serve as starting points for independent investigation, but students do not need to consider these in order to complete the basic experiment. These experiments, as discussed in the context of developing preparatory exercises [34], may be considered to lie somewhere between inquiry and discovery type [35] or between guided and structured inquiry [36, 37], but do not confine students to explore in a particular direction. As well as being introduced to the general philosophy of the laboratories, demonstrators are given instruction on how to guide students in a more Socratic way and the course coordinator monitors their activities. It is noted that demonstrators can have different styles and so care must be taken not to be so prescriptive that individuality is not crushed in the pursuit of instructional compliance.

There is robust evidence that frequent feedback is crucial for students to achieve learning outcomes [38] and that assessments driving these feedback opportunities should emphasise the desired 'skills, knowledge and attributes' [39]. In previous years, the students' laboratory notebooks were discussed individually with a demonstrator every two sessions. These individual discussions were time consuming and meant that demonstrators were rarely on hand to support other students doing the experiments. Furthermore, there was an increase in cohort size from 80 (50) students to over 170 (120) students in the first (second) term over a six year period without a corresponding increase in the number of demonstrators. Hence these discussions were not able to provide the desired level of *timely* feedback expected by students: it was found that 154/174 students (89%) expected demonstrators to be available to answer questions and provide help throughout the entire of the experimental sessions, while only 1% expected demonstrators to be available periodically [9].

In previous years, the course coordinator (PAB) and experienced demonstrators (including MNG and KD) had observed that, despite frequent encouragement and the

suggestions for inquiry contained within the scripts, students were reluctant to explore the experiments beyond the scripts. They were also told that the end of course assessment might require them to explain how they changed their experiments based on observations, but there was no immediate assessment of their actions. One of the main factors making students unwilling to deviate from scripts appeared to be that students were grade focussed, consistent with the findings of [40], and neither the summative nor formative mark schemes directly rewarded exploration. Hence it was proposed that the most efficient way to change student behaviours was to introduce a new element of the laboratory assessment scheme that immediately recognised desired behaviours.

Here, an assessment method, the 'working grade' or 'w-grade' is discussed. It extends and adapts the ideas of [41] with the aim of promoting independent investigation in multi-session experiments. The grade was an evaluation of the type and quality of the work performed by individual students in laboratory sessions. In the work of [41] it was found that such a grade was beneficial towards the acquisition of technical skills, as well as independence and team work. The version discussed here was used in the second term and replaced an earlier trial in the first term of the academic year 2015–2016. Practical work accounted for 85% and a formal report for 15% of the module mark. Within the practical work part, 5% was for a pre-lab activity, 15% for online competency-based tasks, 75% for the end of course assessment, and the remaining 5% (4.25% of the total module mark) was the w-grade.

In the remainder of this paper, we introduce the five aspects of the w-grade (section 2), with examples specific to one of the first term experiments given in the appendix. In section 3, the details of marking are given, followed by a brief discussion of students' marks in section 4. Section 5 contains comments and reflections, and a modified version of the w-grade is outlined in section 6 and conclusions are in section 7.

## 2. Aspects of the w-grade

One of the main considerations when designing the w-grade was the specific types of behaviours that it would aim to encourage [42]. The w-grade was split into five aspects: group work, querying, exploration, attitude and progress. As outlined later in section 3, the first four aspects were marked by two demonstrators independently and the progress mark was assigned at the end of each laboratory session by a single demonstrator.

Students conducted experiments in pairs or occasionally in threes; all aspects were marked individually, although the marks of individual in a group were often similar. It was therefore often difficult to identify imbalances of understanding or involvement unless significant. Marking was done in a similar manner to [41] with a scale 0, 1, 2, 3 in order to recognise and encourage exceptional behaviour which differentiate successfully between those students who really went the furthest and those that did sufficient work to constitute a mark of '2'. As well as the general outline of the w-grade aspects given here, a specific example of one of the experimental scripts, and some of the observed experiment-specific w-grade actions, is given in the appendix.

### 2.1. Group work mark

This aspect considered how the students interacted within their group and was included to encourage students to work as a team, prevent one student from dominating the other(s) and/or prevent other students from being passive. When students experienced problems or obtained unexpected results, they were encouraged to discuss within their group, with other groups in the laboratory and second year students who had encountered the experiments the

**Table 1.** ‘Group work’ mark allocation guide.

---

0	Student either doing most or very little of work.
1	Students sharing workload to a reasonable extent.
2	Both students actively involved in discussing and conducting experiment, equal discussion between groups.
3	Exceptional team work demonstrated.

---

previous year. Discussing work in progress is a key skill for physicists; giving students explicit ‘permission’ to discuss with peers encourages discursive behaviour from early on. Thus key features that demonstrators would be trained to look for include equality of contributions within a group or pair and purposeful, self-organised work leading to an efficient use of the laboratory time. Since demonstrators would typically interact with *c.* 16 pairs of students in a session, the distribution of group work marks would be expected to reflect the relative strengths of the groups performing the particular experiment, with the expectation that most well-functioning groups, engaging in pertinent discussions with peers, would obtain a mark (2).

Many of the students were surprised that we encouraged talking to other groups as they had often been trained to think this was cheating. Indeed, gaining help, both individually and as part of a team, is not ‘cheating’ as long as plagiarism standards are not circumvented. This was presumably because much of their previous practical work was individually assessed based on the achievement of a ‘correct outcome’ after following a rigid laboratory script. This may be, in part, due to the perception of the purpose of practical physics by pre-university teachers [43] and the need to have a simple summative assessment mechanism that takes place during ‘assessment occasions’ [28]. Although encouraging interaction with other groups could sometimes lead to students spending more time asking others for the ‘answers’ than actually doing their own work, this was observed to be rare and instead, genuine discussions on the best approach prevailed. It has also been known for two groups to effectively combine at the data analysis stage, an exemplary example of how rewarding discussion can drastically change attitudes to collaboration. The demonstrators’ outline for mark allocation is given in table 1.

## 2.2. Querying mark

In terms of encouraging students to develop habits associated with critical analysis and understanding of their results, there were two relevant aspects of the w-grade: the ‘querying’ and ‘exploration’ aspects which had slightly different foci.

The ‘querying’ aspect was designed to promote critical thought and penalise students who were clearly seeking the ‘correct’ answer from demonstrators. This generally related to observations during data collection or analysis and how students tried to understand them and the implications for the results or conclusions. The intention was to encourage students to review their understanding in light of unexpected data—or discrepancies in which data that did not quite agree with the given theory. At the lower end (1), the querying mark can be achieved by considering, where relevant, some of the questions marked in the script. At the higher end (3), spontaneous consideration of whether the suggested method is appropriate, as well as a serious attempt to fully understand the details of the experimental set-up and the nuances of repeatability, could be expected. The demonstrators’ marking guidelines are given in table 2.

**Table 2.** ‘Querying’ mark allocation guide.

---

0	Only interested in ‘correct answer/solution’ or no questions asked or answered. Ignoring obvious problems/anomalies.
1	Thoughtful questions asked or thoughtful answers to questions. Realistic consideration of anomalies/problems.
2	Questions exploring nuances—using demonstrator as a ‘sounding board’. Serious attempt(s) to understand anomalies.
3	Exceptional reaction to and insight into the unexpected.

---

**Table 3.** ‘Exploration’ mark allocation guide.

---

0	Accepting all information from script and demonstrator without critical thought.
1	Explored experimental features indicated in script or by demonstrators.
2	Questioned and explored experimental features beyond the limits defined by the script.
3	Unexpected approach to exploring, understanding or conducting the experiment and/or analysis.

---

### 2.3. Exploration mark

The closely related ‘exploration’ aspect aimed to promote independent investigation by rewarding students for going beyond the script and changing the experiment in some considered way. Table 3 gives the demonstrators’ marking guidelines for the ‘exploration’ aspect. The exploration mark is closely linked with the querying mark, it is however distinct since the focus of the latter is on understanding the experiment, both the equipment and the physics, while the exploration mark is about extending the experiment beyond the confines of the script.

This aspect differs from approaches that encourage students to follow a pre-determined, guided path through an experiment [44] as it requires students to directly engage with developments in the experiment, the emergence of unexpected results and personal curiosity as the seeds for further investigation. This extends the training in basic experimental practice and recording techniques which are reinforced and encouraged by other assessment methods that parallel those discussed in [44].

Examples of observed behaviours that scored highly on this aspect included, but were not limited to the following.

- Reworking of the theory to include an additional aspect.
- Significantly changing the basic experimental procedure due to observations.
- Using different or additional equipment to improve the accuracy and precision of results measurement, e.g. using mobile phone cameras to record oscillatory motion.
- Creating new aims in the experiment having obtained a satisfactory result to the initial (defined) aims of the experiment, e.g. investigate the effects of a controllable parameter that only featured tangentially in the basic theory.
- Investigating practical implications of discrepancies between results and theory.
- Asking demonstrators for advice and help based on their specific research experience in their particular research field.

The willingness of students to go a long way beyond the limits of the script was the major success of this aspect which could directly recognise such efforts which previously were not recognised directly in marking criteria.

**Table 4.** ‘Attitude’ mark allocation guide.

0	Student just writing or sitting back and waiting; student arrived late.
1	Student actively involved in work.
2	Student ensuring that partner(s) also involved in work.
3	Exceptional positive attitude to laboratory work.

**Table 5.** Generic achievements or activities by laboratory session for a four-session experiment. Data analysis in particular could occur during any session.

Session 1	Introduction and Theory; Aims and Objectives; Diagrams; Equipment list; Risk assessment; Procedure; Initial data.
Session 2	Modifications to procedure; Data taking; Observations.
Session 3	Repeat/final data; Data analysis.
Session 4	Last data sets and analysis; Conclusions and Reflections.

#### 2.4. Attitude mark

The ‘attitude’ aspect, detailed in table 4, is arguably the most general aspect of the w-grade and has some overlap with the other aspects, particularly group work. The mark scheme does not explicitly consider disruptive students as such students had not been present in previous cohorts. A student who persistently and unnecessarily disturbs or distracts others would obtain a poor mark for both group work and attitude aspects.

Its main aim was to encourage students to develop a conscientious attitude to time keeping, that would serve them well in a professional setting. A good attitude mark was rarely achieved without reasonable performance in the other three areas with the exception that any student who arrived more than ten minutes late to a laboratory session without mitigating circumstances, automatically got a zero. A window of 10 min was allowed for unavoidable public transport delays for which it could be believed that a student had made a reasonable effort to be on time.

This aspect is closely linked with the group work mark, but is a more individual aspect. Thus if students decide that one student will take measurements while the other records the data, the former student is likely to be considered more involved with the experiment, thus earning a higher attitude mark. A good attitude mark would also be awarded for being willing to delve into the nuances of the experiment, perhaps undertaking work between laboratory sessions. This and, to a lesser extent, the group work mark, and possibly even the progress mark, have little variation between different experiments.

#### 2.5. Progress mark

In previous years, inefficient student use of allocated laboratory sessions had been a problem. To prevent accidental loss, laboratory notebooks are kept in the laboratory and although students have access to any data stored on the computers and to the scripts and hence theory, they are not able to spend time between their allocated laboratory sessions undertaking experimental work. Students are discouraged from coming into the laboratory outside of their allocated sessions, but new undergraduates in the first term are granted leniency.

In order to help students make realistic plans for conducting their experiments, each script had a brief summary of expected activities in each session. Although the details were specific to each experiment, they were based on minimum expectations outlined in table 5. In



the middle sessions, where they are taking repeats and modifying the procedure, they would do each activity several times. There was no reason why data analysis should not be done within any session. Table 5 summarises a minimum expectation.

The progress is defined for each session. However, the progress mark for the w-grade was assigned in sessions 2 and 4. Since it was felt to be easily determined from students' laboratory notebooks, it had a no 3 mark; a low progress mark (1 or, rarely 0) was sometimes awarded if no copy of the data was present in the notebook. Some demonstrators found assessing 'progress' hard, citing the subjectivity of this attribute. The tendency was to give a high mark (2) unless the students were clearly not engaged with the work, rather than examining the evidence of the laboratory notebooks. Overall, the evidence at the end of each session should be a good indication that, by the end of the final session, the student will have data based conclusions as the final page(s) of their lab notebook for that experiment.

### 3. Technical aspects and implementation

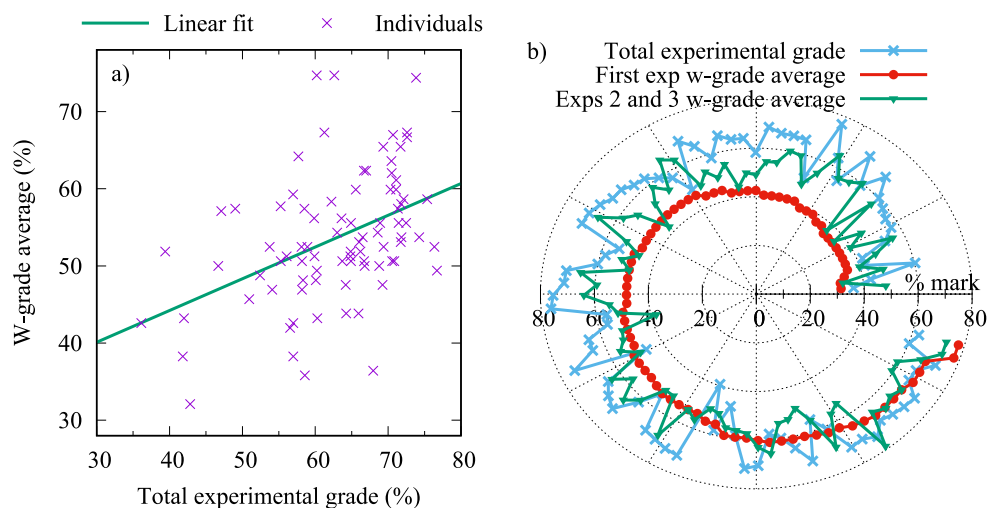
It was considered that the rubric based system outlined above, in conjunction with marks from two demonstrators, would minimise mark allocation variability. This has been used in similar circumstances with positive results [41, 45], and was achieved as discussed below. The timings given worked well, although it took a few sessions before demonstrators could assign marks quickly, especially 'snapshot' marks. Briefing demonstrators did not seem to be sufficient to compensate for their inexperience when marking in laboratory sessions, and there was a tendency to not provide verbal feedback to students.

W-grades were assigned once a week; in the first term this was for every laboratory session, while in the second term implementation discussed here, this was the second and fourth sessions of an experiment. This put less pressure on the demonstrators to upload the marks to UCL's Moodle database immediately and gave the students time to consider what they were doing without fearing that they would get penalised for not actively conducting the experiment or for asking what they thought might be 'stupid' questions.

Each experiment was staffed by two demonstrators assigned to a 'core' cluster of 15–18 students. For the first 90 min of a 3.5 h session each demonstrator worked predominantly with their core cluster, giving a briefing when needed, checking that students were comfortable with the experiment, troubleshooting equipment problems, discussing ideas and answering questions. Then the demonstrators swapped clusters for the next hour where they assigned 'snapshot' marks for all aspects except progress, based on a few minutes observing and discussing with each group of students. They then swapped back to their core cluster for the final hour which allowed the implementation of double marking of all aspects except progress which was assigned by each demonstrator to the students in their core cluster only.

Students were informed of the double marking and also that demonstrators would consider the behaviour observed in both the first and last parts of the session when assigning marks to their core group. Coupled with the expectation of evidence and comments about observations and queries recorded in their laboratory notebooks, this was considered sufficient guard against strong demonstrator-proximity dependence of behaviours. One separate advantage of the double marking was that it encouraged continuous engagement between the demonstrators and students, in accordance with student expectations [9].

The marks of the five individual aspects of the w-grade were recorded on paper forms and then the total w-grade mark for each student uploaded within several days. One issue with this was the students could not see the individual aspect marks; this issue was particularly



**Figure 1.** (a) w-grade marks against total experimental grade (i.e. excluding formal report component of module mark) showing the weak correlation between performance on the w-grade and total experimental mark, and the large variation between individual students. (b) Students are ordered according to their w-grade mark for the first experiment (red, circles), which shows the both the weak correlation to the total experimental grade (blue, crosses) and the increase in w-grade marks from the first to the second two experiments (averaged, green, triangles). The data consists of the marks for the 84 students who had complete w-grade marks for the term (124 on the course).

acute if they received insufficient feedback and explanation of what evidence the demonstrators were using to assign aspect marks. This is discussed in more detail in section 5.

#### 4. Results from students in the 2015–2016 academic year

Obtaining quantitative values related to students' performance is quite difficult because, unlike more constrained introductory physics experiments, it is very hard to control all of the variables or identify from a single quantity whether learning goals have been met. All of our students are individuals with complex motivations and behaviours and different initial attitudes. However, some general trends can be noted in the marks of the 2015–2016 cohort.

Three experiments from the second term of the 2015–2016 academic year are considered. The students had chosen the applied physics course and had been exposed to the earlier version of the w-grade for the two main experiments in the first term. In the first experiment, the w-grade was still relatively new and had been altered compared to what the students had met previously. All students completed the first experiment at the same time, they then completed two further experiments with half of the students doing one experiment first while the other did the other first. They were assessed using the rubrics outlined previously.

From figure 1(a), it can be seen that there is considerable variability among the students. There is a weak correlation between w-grade marks and overall achievement in the laboratory course (excluding the formal report). (Nonlinear least-squares fit to  $y = a + bx$ ,  $y$  is w-grade (%),  $x$  the total lab work (%), with  $a = 27.8$ ,  $b = 0.41$ ; asymptotic standard errors 6.0 (22%) and 0.09 (23%) respectively.) However, it is important to emphasise that there is enormous variation between students, and w-grade values are not a good indicator of achievement.

However, considering that the emphasis of the w-grade is on encouraging students to develop habits of independent inquiry, the change in w-grade marks over the course of a term is more interesting.

In figure 1(b), the average w-grade mark is established for the first experiment and compared with the combined average of the next two experiments. An increase the w-grade is taken as an indication that students gained confidence with the inquiry-style experiments—or at least became more inclined to explore the experiments beyond the confines of the scripts, perhaps more simply as a result of increased understanding that this was accepted behaviour. The students are ordered according to their w-grade mark in the first experiment, and generally improve their w-grades in subsequent experiments. The data consists of the marks for the 84 students of the 124 on the course who completed all of the w-grade assessments during the term (six sessions with w-grade), three had a constant w-grade mark over the term, while the w-grades of 57 students increased (average +10.4%) from the first experiment while only 24 saw a decrease (average −5.3%).

It is important that the w-grade is only a small aspect of the total mark since, although there is some correlation to the final grade, it tends to underestimate students' overall performance, as seen in figure 1(b). This is appropriate for a cohort whose laboratory experience prior to entering first-year laboratories is variable, but, for many students, does not include significant amounts of independent experimental work [9].

## 5. Feedback on initial implementation of the w-grade during 2015–2016

Feedback on the initial w-grade implementation was obtained from students via a free text answer in the 2015–2016 module review questionnaire and from on-going discussions with demonstrators.

### 5.1. From students in 2015–2016 module review

Of the 124 enrolled students, 86 completed the module review questionnaire, of which 40 commented about the w-grade. There were several main themes, with the majority of responses highlighting issues already identified by demonstrators or expressing general dissatisfaction. However, over a quarter of the respondents valued the idea of the w-grade. The major issues fell into two categories.

As noted in [41], students did not appreciate getting low marks for the w-grade, and, despite an outline of the w-grade purpose and aspects being available with the course materials, and a verbal introduction at the beginning of the term, students felt it was unclear how to get top marks. A related problem was with the delayed summative feedback where only the total mark was given and students were unable to see where they had lost marks and therefore could not easily identify and make behavioural changes to improve their mark. Both problems can be addressed quite simply: clearer documentation can be made available to students, and demonstrators should be encouraged to give some sort of verbal feedback at the end of session to either to the entire group or to individuals, explaining w-grade marks or trends. It would also benefit the students if they received their marks for the individual aspects. However, the emphasis of the students' comments was still on how to achieve high marks rather develop skills, returning again to the value students place on marks [46].

### 5.2. Feedback from demonstrators

The subjective nature of the w-grade was the single major concern raised by the demonstrators, echoed in over half (24/40) of the student comments. This could be addressed by improved communication between demonstrators and students about the awarded marks.

As mentioned earlier, many demonstrators were initially uncomfortable with making quick decisions about students' behaviours. Demonstrators were expected to be familiar with the basic method of the experiment and data analysis, and should therefore easily identify and give credit when students did more than the basics. Demonstrators felt that having a far more prescribed set of behaviours to look for would have helped them. Although this was considered before the w-grade was introduced, it was not used since it was felt that it would discourage demonstrators from fully engaging with the experiments. It should be possible to provide some basic guidelines to commonly encountered behaviours and additional training about the those that should be encouraged or discouraged could greatly improve this situation.

The most experienced demonstrators, particularly the course coordinator (PAB) who acts as a 'floating demonstrator' for all sessions and thus interacts with the entire cohort, observed a significant difference in most students' behaviours compared with previous years. Students were far more willing to go beyond the basic experimental procedures outlined in the laboratory script; many asked if they could significantly modify the experiment to investigate their own ideas. Indeed, for the first time, students have asked to use the laboratory facilities to undertake their own studies outside of the physics course programme. This was allowed as it is in line with the overall aims of the UCL teaching laboratories and ethos to encourage research-like activities from early on. The greatest achievement of the w-grade process could be the modification of how students perceive the purpose of a teaching laboratory and what they could do in it. The w-grade ethos, with its emphasis on 'subjective' creativity and inquiry seems to have encouraged these behaviours and given students confidence to conduct independent inquiry.

## 6. Modification to the w-grade for the 2016–2017 academic year

Guided by the students' and demonstrators' comments regarding the first implementation of the w-grade assessment system, and motivated by increased pressures on demonstrators, modifications were made for the 2016–2017 session, as outlined below. Learning from difficulties encountered in 2015–2016, much more care was taken in explaining the purpose and details of the w-grade and marking in general to the students. Further, demonstrators were trained carefully in the use of the now electronic marking.

### 6.1. The new implementation of w-grades

In the past, rubric marking methodologies [47, 48] had been used to help demonstrators give combined summative and formative feedback on students' laboratory notebooks on an individual basis. It was therefore possible, in adapting this system to the increased student numbers, to incorporate the w-grade aspects into a wider assessment and feedback process that largely considered the laboratory notebooks. Inquiry could still be recognised and rewarded, but there was no direct, predefined relation between w-grade and the mark (grade band) awarded after consideration of all elements of interest. After completion of the experiment, at the end of session 4, demonstrators considered the following points.

- Interaction with other groups shown?
- Student showing independent thought?

- Student going beyond the script?
- Student fully involved in work?
- Sufficient progress in time allowed?

The demonstrators asked students to show, using their laboratory notebooks, how they achieved these competencies. There were three levels that could be achieved, with descriptive definitions: consistent, partial and lacking evidence, which could be easily transformed into marks, a possibility that was not used in the implementation reported here. Although this simplified method appears to be more subjective since particular behaviours are not specified, it seemed to be less ambiguous and therefore easier for the demonstrators to implement and more understandable for the students. The demonstrators' decisions were entered directly into an online form, speeding up the marking and feedback.

In addition, students received pre-written feedback statements if they were not consistently demonstrating a particular aspect. This helped to standardise basic responses to students and helped demonstrators understand what they needed to look for in the students' behaviours. The automatic feedback comments were designed to help students understand what demonstrators were expecting of them and provide some direction for improvement. The feedback comments, given in the same order as the questions above, were as follows.

- Interacting with other groups is essential for experimentalists. You must do this without being told to do so. Compare ideas, thoughts, data and results. Theorists do this too ... and at the end: 'collaboration is not cheating, if you are worried about plagiarism, make a note of who you discussed ideas with'?
- You need to show that you are trying to resolve problems yourself rather than relying on demonstrators to do this for you.
- Pushing the boundaries of science is what physicists do. We need you to do this as a standard behaviour.
- You must be fully engaged with the work. You need to be able to demonstrate this to demonstrators.
- You must use your time efficiently. Failure to do so will reduce your experimental effectiveness. Plan ahead.

These feedback notes were designed to help students understand not only what demonstrators were expecting of them, but also be supportive for students who were still learning to explore beyond the scripts.

## 6.2. Feedback from students in 2016–2017 module review

One hundred and seventeen students took the laboratory course in 2016–2017, of which 46 completed the end of module questionnaire, and 21 gave specific comments on the w-grade component of the assessment process. There were still some concerns about fairness, but the lower weighting of the w-grade component in this combined assessment scheme means that the impact on overall (graded) performance is minimal. Some examples of comments follow.

- *Encouraged to go beyond the laboratory script and collaborate with other groups.*
- *It is good that the marking is done while discussing with the students and that oral part is the more important part ...*
- *Very useful—motivated you to find extra things during the experiment that otherwise you may have ignored.*
- *Useful, promotes going on further than just the base experiment.*

It seems that the students preferred the w-grade as part of a wider assessment of their performance. It is not unusual that they seemed to favour an individual verbal assessment that gives feedback just as they have completed a task [49]. With this variant of the w-grade, students also seemed to be moving further away from a purely mark optimisation approach to laboratory work. There are, however, several significant weaknesses to this variant. In particular, the fact that the assessment now occurs at the end of each experiment, means that there is less emphasis on habit-forming, efficiently reinforced by frequent formative feedback [38] and the scope of the discussion is also not nearly as comprehensive as in the full version.

## 7. Conclusions

The introduction of a w-grade to a large cohort first-year undergraduate physics laboratory course has been discussed. The assessment method aims to encourage five key behaviours of effective group work, querying, exploration, positive attitude and time-management (progress) using a simple 0–3 marking format for each of the five aspects with double marking implemented to reduce variations between demonstrators [41]. This is partly based on the idea presented in [41], which has been adapted and extended to multiple session experiments, and its flexibility demonstrated. An analysis of the w-grade marks across the second term indicates that the w-grade method can change students' behaviours regarding these 'soft' experimental skills. As well as the full implementation of the w-grade discussed in detail, a simpler variant which requires less demonstrator engagement and is more palatable to students, but suffers from reducing the assessment to an end of activity review rather than continuous on-going feedback, has also been summarised. By presenting the two versions together, the flexibility and potential of the w-grade concept for laboratory courses with different aims and resource pressures should be apparent.

We suggest that the w-grade can be a means of directly rewarding inquiry and boosting self-esteem in an undergraduate laboratory environment by promoting willingness to expand their approach to experimental physics and engage with potentially 'wrong' ideas and 'answers'. Most difficulties experienced involved the feedback system which initially prevented students from knowing their individual aspect marks and explanation of how they had shown insufficient evidence to gain desirable marks. Changing the marking mechanism and simplifying the criteria made the w-grade more accessible to both students and demonstrators. It also had the added benefit that the marks reported directly on observed behaviours. Simplicity and clarity of the marking scheme was crucial for efficient implementation by demonstrators and for students to appreciate the purpose of the assessment.

## Acknowledgments

Demonstrators and students in UCL's first-year undergraduate physics laboratories 2015–2016; 2016–2017. We thank D Armoogum for useful discussions and help with the rubric implementation in 2016–2017, and J Grozier for his on-going support and relentless fine tuning of laboratory scripts.

## ORCID iDs

K Dunnett  <https://orcid.org/0000-0003-2486-5802>

M N Gorman  <https://orcid.org/0000-0002-8437-1284>

P A Bartlett  <https://orcid.org/0000-0003-4323-1485>

## References

- [1] Holmes N G, Olsen J, Thomas J L and Wieman C E 2017 Value added or misattributed? a multi-institution study on the educational benefit of labs for reinforcing physics content. *Phys. Rev. Phys. Educ. Res.* **13**
- [2] Holmes N G and Wieman C E 2018 Introductory physics labs: we can do better *Phys. Today* **71** 38–45
- [3] Hofstein A and Lunetta V N 2004 The laboratory in science education: Foundations for the twenty-first century *Sci. Educ.* **88** 28–54
- [4] Wieman C 2015 Comparative cognitive task analyses of experimental science and instructional laboratory courses *Phys. Teach.* **53** 349–51
- [5] Hofstein A and Lunetta V N 1982 The role of the laboratory in science teaching: Neglected aspects of research *Rev. Educ. Res.* **52** 201–17
- [6] Lunetta V N and Tamir P 1979 Matching lab activities with teaching goals *Sci. Teach.* **46** 01
- [7] Fisher D, Harrison A, Henderson D and Hofstein A 1998 Laboratory learning environments and practical tasks in senior secondary science classes *Res. Sci. Educ.* **28** 353–63
- [8] Meltzer D E and Otero V K 2015 A brief history of physics education in the United States *Am. J. Phys.* **83** 447–58
- [9] Dunnett K and Bartlett P A Intake questionnaires, UCL physics department. Since 2015–2016, students have been asked about their prior experience, expectations and current perceptions of the purpose of laboratories, in their very first undergraduate laboratory session. Apart from the purpose of laboratories (introduced 2016–2017), the data referred to is for the 2015–2016 academic year, but all years are similar, 2015–2018
- [10] Marx R W, Freeman J G, Krajcik J S and Blumenfeld P C 1998 Professional development of science teachers *International Handbook of Science Education* ed B J Fraser and K G Tobin (Dordrecht: Kluwer) pp 667–80
- [11] Ryan B J 2014 Graduate teaching assistants; critical colleagues or casual components in the undergraduate laboratory learning? an exploration of the role of the postgraduate teacher in the sciences *Eur. J. Sci. Math. Educ.* **2** 98–105
- [12] Howitt S, Wilson A, Wilson K and Roberts P 2010 Please remember we are not all brilliant: undergraduates experiences of an elite, research-intensive degree at a research-intensive university *Higher Educ. Res. Dev.* **29** 405–20
- [13] Laws P 1991 Workshop physics: Learning introductory physics by doing it *Change: Mag. Higher Learn.* **23** 20–7
- [14] Cummings K, Marx J, Thornton R and Kuhl D 1999 Evaluating innovation in studio physics *Am. J. Phys.* **67** S38–44
- [15] Beichner R J, Saul J M, Abbott D S, Morse J J, Deardorff D, Allain R J, Bonham S W, Dancy M H and Risley J S 2007 The student-centered activities for large enrollment undergraduate programs (SCALE-UP) project *Research-Based Reform of University Physics* vol 1 (College Park, MD: American Association of Physics Teachers) pp 2–39
- [16] Docktor J L and Mestre J P 2014 Synthesis of discipline-based education research in physics *Phys. Rev. Spec. Top. - Phys. Educ. Res.* **10** 020119
- [17] Laws P W, Willis M C and Sokoloff D R 2015 Workshop physics and related curricula: A 25-year history of collaborative learning enhanced by computer tools for observation and analysis *Phys. Teach.* **53** 401–6
- [18] Strubbe L E, Ives J, Holmes N G, Bonn D A and Sumah N K 2016 Developing student attitudes in the first-year physics laboratory *Physics Education Research Conf. 2016, PER Conf. (Sacramento, CA, 20–21 July)* pp 340–3
- [19] Etkina E, Murthy S and Zou X 2006 Using introductory labs to engage students in experimental design *Am. J. Phys.* **74** 979–86
- [20] Etkina E, Karelina A, Murthy S and Ruibal-Villasenor M 2009 Using action research to improve learning and formative assessment to conduct research *Phys. Rev. Spec. Top. - Phys. Educ. Res.* **5** 010109
- [21] Hmelo-Silver C E, O'Donnell A M and Erkens G 2002 Design principles for scaffolding technology-based inquiry *Rutgers Invitational Symposium on Education; Collaborative Learning, Reasoning, and Technology (New Brunswick, NJ)* (Mahwah, NJ: Erlbaum)
- [22] Kempa R F and Martín-Díaz M 1990 Motivational traits and preferences for different instructional modes in science: I. Students' motivational traits *Int. J. Sci. Educ.* **12** 195–203

- [23] Ames C and Archer J 1988 Achievement goals in the classroom: Students' learning strategies and motivation processes *J. Educ. Psychol.* **80** 260–7
- [24] Dweck C S 1986 Motivational processes affecting learning *Am. Psychol.* **41** 1040–8
- [25] Swain J R L 1991 The nature and assessment of scientific explorations in the classroom *Sch. Sci. Rev.* **72** 65–77
- [26] Blumenfeld P C 1992 Classroom learning and motivation: clarifying and expanding goal theory *J. Educ. Psychol.* **84** 272–81
- [27] Newman R S and Schwager M T 1995 Students' help seeking during problem solving: effects of grade, goal, and prior achievement *Am. Educ. Res. J.* **32** 352–76
- [28] Hodson D 1992 Assessment of practical work *Sci. Educ.* **1** 115–44
- [29] Morss K and Murray R 2005 *Teaching at University: A Guide for Postgraduates and Researchers* (London: Sage)
- [30] Chang H P and Lederman N G 1994 The effect of levels of cooperation within physical science laboratory groups on physical science achievement *J. Res. Sci. Teach.* **31** 167–81
- [31] Hodson D 1993 Re-thinking old ways: towards a more critical approach to practical work in school science *Stud. Sci. Educ.* **22** 85–142
- [32] Wilkinson J W and Ward M 1997 The purpose and perceived effectiveness of laboratory work in secondary schools *Aust. Sci. Teach. J.* **43** 49–55
- [33] MacIsaac D 2015 Report: Aapt recommendations for the undergraduate physics laboratory curriculum *Phys. Teach.* **53** 253–253
- [34] Dunnett K and Bartlett P A 2017 Asking the next generation: the implementation of pre-university students' ideas about physics laboratory preparation exercises *Phys. Educ.* **53** 015016
- [35] Domin D S 1999 A review of laboratory instruction styles *J. Chem. Educ.* **76** 543
- [36] Bell R L, Smetana L and Binns I 2005 Simplifying inquiry instruction *Sci. Teach.* **72** 30–3
- [37] Buck L B, Bretz S L and Towns M H 2008 Characterizing the level of inquiry in the undergraduate laboratory *J. Coll. Sci. Teach.* **38** 52–8
- [38] Black P and Wiliam D 1998 Assessment and classroom learning *Assess. Educ.: Princ., Policy Pract.* **5** 7–74
- [39] Crooks T J 1988 The impact of classroom evaluation practices on students *Rev. Educ. Res.* **58** 438–81
- [40] Frost G and Connolly M 2016 Is fine tuning possible with grade-focused students? *Collected Essays Learn. Teach.* **9** 147–54
- [41] Lippi G L 2015 Improving students' lab practices: the performance grade arXiv:1503.07182v1 physics.ed-ph
- [42] Wilcox B R and Lewandowski H J 2017 Developing skills versus reinforcing concepts in physics labs: Insight from a survey of students' beliefs about experimental physics *Phys. Rev. Phys. Educ. Res.* **13** 010108
- [43] Hirvonen P E and Viiri J 2002 Physics Student Teachers' ideas about the objectives of practical work *Sci. Educ.* **11** 305–16
- [44] Etkina E and Murthy S 2006 Using introductory labs to engage students in experimental design *Am. J. Phys.* (<https://doi.org/10.1119/1.2238885>)
- [45] Lippi G L 2014 A priori analysis: an application to the estimate of the uncertainty in course grades *Eur. J. Phys.* **35** 045012
- [46] Gibbs G 2006 How assessment frames student learning *Innovative Assessment in Higher Education* (London: Routledge) p 23
- [47] Bartlett P A 2014 Data retrieval tests and task book training: an alternative route to laboratory assessment *Paper presented at the UCL Teaching and Learning Conf. (London)*
- [48] Etkina E, Van Heuvelen A, White-Brahmia S, Brookes D T, Gentile M, Murthy S, Rosengrant D and Warren A 2006 Scientific abilities and their assessment *Phys. Rev. Spec. Top. - Phys. Educ. Res.* **2** 020103
- [49] O'Donovan B, Rust C and Price M 2016 A scholarly approach to solving the feedback dilemma in practice *Assess. Eval. Higher Educ.* **41** 938–49