# Grade distribution digests: A novel tool to enhance teaching and student learning in laboratory practicals.

Running title: Grade distribution digests

*Authors*

Peter G. Arthur, Reza Zareie, Paul Kirkwood, Martha Ludwig, Paul V. Attwood

*Abstract*

Assessment is a central component of course curriculums and is used to certify student learning, but it can also be used as a tool to improve teaching and learning. Many laboratory courses are structured such that there is only a grade for a particular laboratory, which limits the insights that can be gained in student learning. We developed a laboratory program that incorporates assessments designed to probe student understanding of different components of the individual modules making up the program. The challenge was to analyse and present grades from these assessment tasks in a format that was readily interpretable by academics. We show that a simplified synthesis of grade distributions (grade distribution digests) provides sufficient information to make decisions about changes in course components. The main feature of the digests is its data visualization approach, where student grades for individual laboratory practicals, individual assessment tasks or individual assessment items are graphically presented as an overall average grade, an average top quartile grade and an average bottom quartile grade, and relative averages across all assessments. This ability to visualize student grades in variety of contexts enables academics with many other demands on their time (e.g. research and administration) to more efficiently identify ways to improve teaching delivery and learning outcomes. Examples are presented of the use of such data to identify and improve deficiencies in both student skills and teaching practice, resulting in improved learning outcomes.

## Introduction

In Australia and other countries, systemic issues have been identified that hinder learning in science courses [1, 2]. For example, the passive learning style of laboratory practicals has been highlighted as one reason for poor learning outcomes [2]. As a consequence, innovative practices for teaching scientific inquiry skills have been developed to engage students and enhance their scientific inquiry skills [1, 3-5]. Examples of scientific inquiry skills include ability to design experiments; formulate hypotheses; generate, analyse, and interpret data; conduct statistical tests; draw conclusions; critically appraise information; collaborate; and communicate [6-8]. However, evaluating the effectiveness of student acquisition of these skills in teaching practice is challenging [5].

In 26 case studies of laboratory practicals evaluated by Elliott *et al* [5]: "there was little evidence to show that a particular teaching approach achieved its goal". Evidence for effectiveness of teaching innovations included feedback from focus groups and questionnaire data including: "More students agreeing that a task helped them improve" and "Students agreeing that a task was relevant or had given them insight into how research is conducted". Perception by educators was also used as evidence of effectiveness and included judgements of increased competency, improved assigned grades or the course becoming increasing popular. However, these types of broadly defined evaluations do not usually provide sufficient detail to establish whether students have successfully acquired particular scientific inquiry skills. [5, 9, 10].

Assessment is a central component of course curriculums and is used to certify student learning, but it can also be used as a tool to improve teaching and learning [11-14]. Assessments can be used to evaluate student acquisition of scientific inquiry skills for specific course components and the data can then be used to guide decisions about how to modify the course to improve student learning [10] [15]. However, in many laboratory courses there is only a single grade for a particular laboratory or group of laboratories, which does not provide sufficient information about student acquisition and application of scientific inquiry skills.

An alternative approach is to incorporate more detailed assessments to evaluate student acquisition of scientific inquiry skills in different aspects of the laboratory course [12, 16]. We therefore developed a structured laboratory program (Laboratory Modules in Contemporary Technologies, LMCT) designed to develop scientific skills and to engage students, using a teaching model suited to large groups [6]. Multiple assessments form the basis of the LMCT program, with the questions designed to assess student acquisition of scientific inquiry skills for the different components in each laboratory (summarized in Table 1). The assessment of each laboratory practical fell into two main categories: pre-laboratory (pre-lab) and post-laboratory (post-lab). The pre-lab multiple-choice question (MCQ) quiz was designed to test student knowledge of the technique and its specific application in the up-coming practical, based on material provided in an online lecture. The aim here was to ensure the students were prepared before they came into the laboratory. The post-lab assessment comprised both MCQ and a short answer question (SAQ) quizzes. The aim in these assessments was to test the students' understanding of the experiment they had performed in the laboratory and their skills in data calculation and presentation, use of statistical and graphical analyses of data and data interpretation. This assessment was open-book to minimize the testing of memorized material and focus on assessment of acquired skills. For a detailed description of the LMCT program, including the assessment tasks see Arthur *et al* [6].

Our objective was to use grades from the assessments to make decisions about changes in course components to optimize student acquisition of scientific inquiry skills. As part of this process, we were interested in a tool that could be used to compare different aspects of the laboratory course. Comparisons of interest were grades between laboratories, grades between course components (e.g. pre-lab and post-lab assessments); and grades within a course component (e.g. the 4 questions in a pre-lab quiz). The challenge was to analyze and present grades in a format that was readily interpretable by academics [17]. In addressing this challenge, we developed a simplified synthesis of grade distributions, which we call grade distribution digests (GDDs). In this manuscript, we first describe the rationale for the GDD concept, and then show how GDDs can be used to generate quantitative evidence to make informed decisions about course components to enhance student learning.

**Rationale for the Grade Distribution Digest concept**

Multiple assessments form the basis of the LMCT program, but this generates a considerable number of grades, which we found cumbersome to use in making decisions about course content. For example, in the second semester laboratory program there are 51 separate grades for each student, which are comprised of 20 pre-laboratory MCQ, 20 post-laboratory MCQ, five written answers, five results sheets and one presentation completed during the class (Table 1). For a class of 100 students, this represents 5100 separate grades.

The challenge was to analyze and present grades in a format that was readily interpretable by academics, so that targeted decisions about changes in course components could be made. There are a number of approaches to analyze and present student grades such as: averages with measures of error, grade distributions, individual item analysis, or more sophisticated statistical analyses such as the Rasch model [17, 18]. The Rasch model or individual item analysis could be used to provide information about the difficulty of each question. However, as noted by Crisp and Palmer [17], instructors are not usually familiar with these tools and their results can be misinterpreted. As an alternative, we examined whether using average grades or grade distributions could inform decisions about course components. We found these measures did not provide sufficient information (averages) or were not readily interpretable (grade distributions) to come to useful conclusions about student understanding of specific course components.

In seeking a solution to the grade interpretation challenge, we focused on two desirable outcomes that have been identified in setting individual questions [18]. The first outcome is that motivated and capable students who have acquired the requisite scientific inquiry skills will be able to answer the questions correctly (see section “Using GDDs to change course components” for examples). The second outcome is that of grade discrimination, so the less motivated or less capable students receive lower grades. For these two outcomes, we modelled four scenarios of grade distributions in response to multiple questions (Fig. 1).

Scenario 1

An example of a grade distribution whereby capable students are able to answer most of the question(s) while less capable or less motived students received lower grades (Fig. 1A). This grade distribution was modeled using a standard deviation calculated from course grades aggregated over 5 years for a second year, one semester biochemistry course. Course grades were calculated from theory (mid-semester and end of semester) exams and laboratory grades.

Scenario 2

An example of a grade distribution that is skewed towards lower grades (marks out of 100) (Fig. 1B). This would indicate that the questions were too challenging for capable students. This distribution could reflect poorly worded questions or deeper issues such as insufficient training or inadequate background theory (see section ‘Using GDDs to change course components’ for examples).

Scenario 3

An example of a grade distribution that is skewed towards high grades (marks out of 100) (Fig. 1C). This distribution would indicate that the questions were not sufficiently challenging, resulting in a lack of discrimination of student capabilities.

Scenario 4

An example of a grade distribution that is bimodal such that some students are able to answer the questions, but other students found the questions too difficult (Fig. 1D). This distribution could occur when, for example, a cohort of students has insufficient prerequisite background.

FIGURE 1 NEAR HERE

The four scenarios of grade distributions can be more simply represented by calculating a mean grade for the bottom quartile of grades and the top quartile of grades (Fig. 1). In practice, we have found that including average grades, an idea with which most academics are familiar [17], assists in understanding the concept of the top and bottom quartiles. This simplified synthesis of grade distribution, which we refer to as a grade distribution digest (GDD), allows the four example scenarios of grade distribution (Fig. 1) to be easily distinguished.

A grade distribution that differs from a target (or reference) grade distribution (Fig 1A) identifies weakness in teaching delivery or student acquisition of scientific inquiry skills, which then can be corrected. As the example shows, the type of grade distribution (Scenarios 2-4, Fig. 1) can be readily identified if there is a target grade for the bottom and top quartiles (Fig. 1). Because of differences in opinions about suitable grade distributions for courses, target grades for bottom and top quartiles can be expected to differ between courses. For our laboratory course, we used grades collated from the course for the previous five years, presented in a format typically used by our School (School of Molecular Sciences, The University of Western Australia) to represent grades (Fig. 2A). From these data, we calculated the target grades for the bottom quartile and top quartile of students, as well as the class average (Fig. 2B).

FIGURE 2 NEAR HERE

**Presentation of data**

Teaching, research and administrative workloads put pressure on the time available for course improvement [17, 19-22]. Many staff who are involved in designing, running and administering laboratory teaching practicals have numerous demands on their time as they are also active researchers and have additional school-level and university-level administrative duties. Crisp and Palmer [17], suggested visual engagement would allow time-poor university staff to quickly determine the salient issues for a particular assessment. Visualization is recognized as a useful medium for examining, understanding and transmitting information [23, 24] and has been used to enhance interpretation and understanding of exam results [22, 25]. To visually report the GDDs (Fig. 2B), we therefore used the principles of data visualization described by Iliinsky and Steele [23], the key cornerstone of which is the idea of the interplays within the designer-reader-data trinity. Thus a key to effective and persuasive communication of data is the consideration of the users knowledge and experience. As there are also different academic preferences for the way data are presented [17], we also report data numerically (Fig. 2C).

**Analysis of laboratory grades using Grade Distribution Digests.**

The premise of GDDs is that by checking two aspects of grade distributions it is possible to identify weaknesses in course delivery and assessment. The first check, based on grades for the top quartile of students, is that motivated and capable students will have the requisite skills and be able to answer the question(s) correctly. Questions that are too challenging for capable students may reflect poorly worded questions or deeper issues such as inadequate teaching, training, or knowledge of background theory. The second check, based on grades for the bottom quartile of students, is that of grade discrimination to identify less motivated or capable students. A lack of grade discrimination could indicate that questions were not sufficiently challenging to capable students and could also create the perception that instructors are more effective than they actually are [14].

We used the GDD approach to evaluate student grades for course components for five laboratories run over the course of a semester (Fig. 3). Each laboratory consisted of four assessed components: (1) Preparation, using an online lecture to describe the theory of a technology. This was assessed using an online quiz of four MCQ (Fig. 3, Prelab MCQ); (2) A laboratory session involving “hands on” application of the technology. During the laboratory session, students generated data and then answered questions about their data. This was graded during the laboratory by a demonstrator (usually a graduate student, also referred to as a teaching assistant) responsible for overseeing a group of students undertaking the laboratory (Fig. 3, Demo); (3) A post-laboratory session involving student-driven presentations of the technology. An open book test was then used to assess student understanding and consisted of four MCQ (Fig. 3, Postlab MCQ) and (4) a SAQ (Fig. 3, Postlab SA). For the four assessed components, a GDD was generated for the semester (Fig. 3). Student grades for the same component across all laboratories can be evaluated by a vertical comparison, with horizontal comparisons used to evaluate different components within the same laboratory (Fig. 3).

FIGURE 3 NEAR HERE

The above overview of the laboratory course makes it is possible to quickly identify weaknesses in aspects of course delivery and assessment. The data were used in two ways: (1) to examine student grades for the same component across several laboratories and (2) to examine student grades in various components of a single laboratory. From inspection of the GDD for Laboratory 1 (Fig. 3), it is evident there is a lack of discrimination for the demonstrator's grade (Demo). The lack of discrimination was not a consequence of a particular fault with Laboratory 1, as it was evident across all laboratories. This indicates a systemic issue with demonstrator grading. To explore this further, grades across the five laboratories were aggregated by course component and compared to target reference grades (Fig. 4A). It was apparent that demonstrator grades were substantially higher than reference grade targets.

In debating the high grades and lack of grading discrimination, we considered the role of the demonstrator in the class. In our original laboratory course, which was replaced by the LMCT course, the demonstrators were responsible for a substantial proportion of the grades for the laboratory practicals. However, there were ongoing issues in ensuring grading equity between demonstrators. In addition, the grade distributions were skewed towards high grades in the original laboratory course, which indicated a lack of grade discrimination (Fig. 4B). In the LMCT course, we changed the demonstrator's primary role to that of engaging with students, rather than focusing on student assessment. Our rationale was that by removing the pressure to assess, the demonstrator could focus on interacting with students and assist in developing their scientific skills. Given this concept, we decided to accept the lack of grade discrimination, but made the grade only a minor (10%) component of the total grades for the laboratory. The demonstrator grade can be considered to be a compliance grade such that the demonstrator still provided input into student assessment and students were still aware of the need to perform satisfactorily.

FIGURE 4 NEAR HERE

This outcome for demonstrator grading also illustrates two other aspects of teaching practices related to the use of GDDs. First, it stimulated us to engage in reflective teaching practice, an important component in developing effective teaching [19], by considering the role of demonstrators in the laboratory. Second, we accepted the lack of grading discrimination for demonstrators, which indicates that a lack of grade discrimination does not always have to result in modifications to the laboratory course.

**Using Grade Distribution Digests to change course components.**

Through the use of GDDs, weaknesses in course delivery were readily identified, and changes could be implemented. The effectiveness of these changes was then measured. The following examples illustrate this use of GDDs.

Example 1 – course improvement.

In this example, for the third laboratory (Polymerase Chain Reaction) of a first semester course (Table 1), the grade distribution was skewed low for the top and bottom quartiles of the student cohort (Fig. 5A, year 1). An analysis of the different components of the laboratory, showed that students struggled to answer the post-laboratory SAQ, which was a calculation style question (Fig. 5B). As the top quartile of student grades was also substantially lower than average for this question compared with the SAQ of all laboratories, this indicated a structural issue. The particular question involved calculations of concentration. The academic responsible for this laboratory realized that the students may not have had the expected level of quantitative skills to answer the question. As a consequence, the academic provided additional written guidance, including detailed examples, in the next iteration of the laboratory in the following year (See Supplemental 1). Additionally, emphasis on quantitative analysis was increased by including additional formative calculation questions during the hands-on laboratory session and in the online lecture for this laboratory. In the following year, the mean grade for the top quartile of students increased from 42% to 88% (Fig. 5C, year 2).

FIGURE 5 NEAR HERE

Example 2.

In this example, academic responsible for laboratory 1 (see Table 1) noted that student grades for the post-laboratory MCQs were lower than the average for all laboratories (Fig. 6A, year 1). This was not a consequence of an overall poor understanding of the laboratory as the student grades were higher for the post-laboratory SAQ (Fig. 6A). A breakdown of the MCQ grades showed that this was not a systemic issue for all post-laboratory MCQ (Fig. 6B). Rather it was a single question (Fig. 6B, MCQ4) that was poorly answered with respect to others. A review of the laboratory indicated that the poor response was possibly a consequence of a lack of background information. Additional information was added to relevant section of the laboratory (73 words increased to 155) (see supplement 2). In the following year, without change to the question itself, there was almost a doubling in the percentage of students able to answer the question correctly (Fig. 6C, year 2).

FIGURE 6 NEAR HERE

**Summary**

Student grades are a valuable resource to improve teaching and learning. However, as identified by Crisp and Palmer [17], non-specialist education staff have difficulties in analyzing and using assessment data. Crisp and Palmer [17] suggested that academics require reports that can be quickly interpreted so that issues impacting on student performance could be readily identified. Our simplified synthesis of grade distributions (GDDs) can be used to monitor student performance in laboratories. While this mode of presentation is not as detailed and comprehensive as approaches to analyzing grades, such as item analysis [17, 18] [26], it can be readily explained to, and understood by, non-education specialist academics.

As shown by the examples, we were able use information from GDDs to identify specific weaknesses in aspects of the laboratory program and then change these aspects to improve student understanding and learning. In subsequent iterations of the laboratory program, GDDs were used to assess whether the modifications met these goals. This use of GDDs enables an iterative curriculum development cycle between instruction, assessment and learning outcomes and has been recommended as a means of improving teaching and enhancing the student experience [13, 19, 27, 28] [10].

Standard statistical methods associated with analyzing the validity and reliability of an assessment are focused on individual questions [26]. In contrast, grades can be aggregated in different ways for GDDs, for example, GDDs can be generated from grades for a single SAQ, grades involving multiple questions (e.g. prelaboratory questions), and for final grades for the laboratory. As a consequence, GDDs can be used to provide a broader perspective on student performance between different laboratories or laboratory components.

*Acknowledgment*

This work was supported by an Improved Student Learning Grant from The University of Western Australia to PGA, ML and PVA. The authors have no conflict of interest to declare.

**TABLE 1.**

Assessable items in two one-semester laboratory courses. Assessable items are grouped horizontally by laboratory component and vertically by laboratory number. Prelab MCQ, prelaboratory multiple-choice questions; Demo, demonstrator or teaching assistant grade; Postlab MCQ, post-laboratory multiple-choice questions; Postlab SA, post-laboratory short answer question. All multiple-choice questions had five possible answers. For some items (Prelab MCQ, Postlab MCQ and Postlab SA) there were randomized alternate questions, but these were treated as a single assessed item in the analysis. Most laboratories incorporated one session of experimental work except two laboratories incorporated two sessions of experimental work over two weeks (indicated by *). Further details about the laboratory course are described in [6].

| 1st semester LMCT program | Number of assessed items | | | | |
|---|---|---|---|---|---|
| Laboratory number and title | Prelab MCQ | Demo | Postlab MCQ | Postlab SA | Total |
| 1 Homogenization & Centrifugation | 4 | 1 | 4 | 1 | 10 |
| 2* DNA Hybridisation | 4 | 1 | 4 | 1 | 10 |
| 3* Polymerase Chain Reaction | 4 | 1 | 4 | 1 | 10 |
| 4 Enzyme Kinetics | 4 | 1 | 4 | 1 | 10 |
| Total | 16 | 4 | 16 | 4 | 40 |
| Total +1 presentation | | | | | 41 |

| 2nd semester LMCT program | Number of assessed items | | | | |
|---|---|---|---|---|---|
| Laboratory number and title | Prelab MCQ | Demo | Postlab MCQ | Postlab SA | Total |
| 1 Protein Purification | 4 | 1 | 4 | 1 | 10 |
| 2 Working with Fluorescence | 4 | 1 | 4 | 1 | 10 |
| 3 Electrophoresis I SDS-PAGE | 4 | 1 | 4 | 1 | 10 |
| 4 Electrophoresis II Western Blotting | 4 | 1 | 4 | 1 | 10 |
| 5 Cell Signalling | 4 | 1 | 4 | 1 | 10 |
| Total | 20 | 5 | 20 | 5 | 50 |
| Total + 1 presentation | | | | | 51 |

**FIGURES.**

Figure 1. Example grade distributions.

Grades are given as marks out of 100. Models of grade distributions were generated to illustrate grade distributions that were (Ai) on target, (Bi) skewed low, (Ci) skewed high or (Di) bimodal. Grade distributions are shown in a format typically used to represent grades (Aii, Bii, Cii, Dii). The visualization format (Aiii, Biii, Ciii, Diii) shows the average grade for the class (blue horizontal column), mean grade for the top quartile of students (green horizontal column) and the mean grade for the bottom quartile of student (red horizontal column). Black vertical bars (Biii, Ciii, Diii) are reference indicators, showing target grades for the target distribution (from Aiii).

Figure 2. Course grade distributions.

Grades are given as marks out of 100. Course grade distributions (theory plus laboratory) were averaged over five years and 95% confidence intervals were calculated. (A) Grade distributions in a format typically used to represent grades. (B) Course grade distribution expressed in the GDD format. (C) Visual representation of GDD format. For (A) and (C), error bars are 95% confidence intervals.

Figure 3. Grade Distribution Digest for a second semester laboratory course.

Grades are given as percentages. (A) Visual representation of data scaled by percentage. Assessable items are grouped by laboratory and by laboratory component as described in Table 1. To assist in comparisons, vertical lines show mean grades for the particular component of the five laboratories for bottom quartile of students (red lines), average of all students (blue lines) and top quartile of students (green lines). (B) Numerical presentation of data shown in (A) with 95% confidence intervals in brackets.

Figure 4. Summary Grade Distribution Digest for two different versions of a first semester laboratory course.

Grades are given as percentages. (A) Grades were aggregated by course component from five laboratories in the LMCT course as described in Table 1. The table shows the numerical presentation of data with 95% confidence intervals in brackets. (B) In the earlier version of the laboratory course, all grades were provided by demonstrators (teaching assistants). Confidence intervals (95%) are presented visually as error bars and numerically with 95% intervals bracketed. To show the extent the grades differed from the target reference grades (taken from Fig. 2), vertical lines show target reference grades of the bottom quartile of students (red lines), average of all students (blue lines) and top quartile of students (green lines).

Figure 5. Identifying and correcting poor grade distributions – Example 1.

Grades are given as percentages. (A) Grade Distribution Digest for Laboratory 3, year 1. (B) Expanded data set for Laboratory 3, year 1. (C) GDD for Laboratory 3 short answer question following modification (year 2) to the laboratory (see main text) compared with the previous year (year 1). To show the extent to which grades differed in this laboratory from the average for all laboratories, vertical lines show mean grades for the five laboratories for bottom quartile of students (red lines), average of all students (blue lines) and top quartile of students (green lines).

Figure 6. Identifying and correcting poor grade distributions – Example 2.

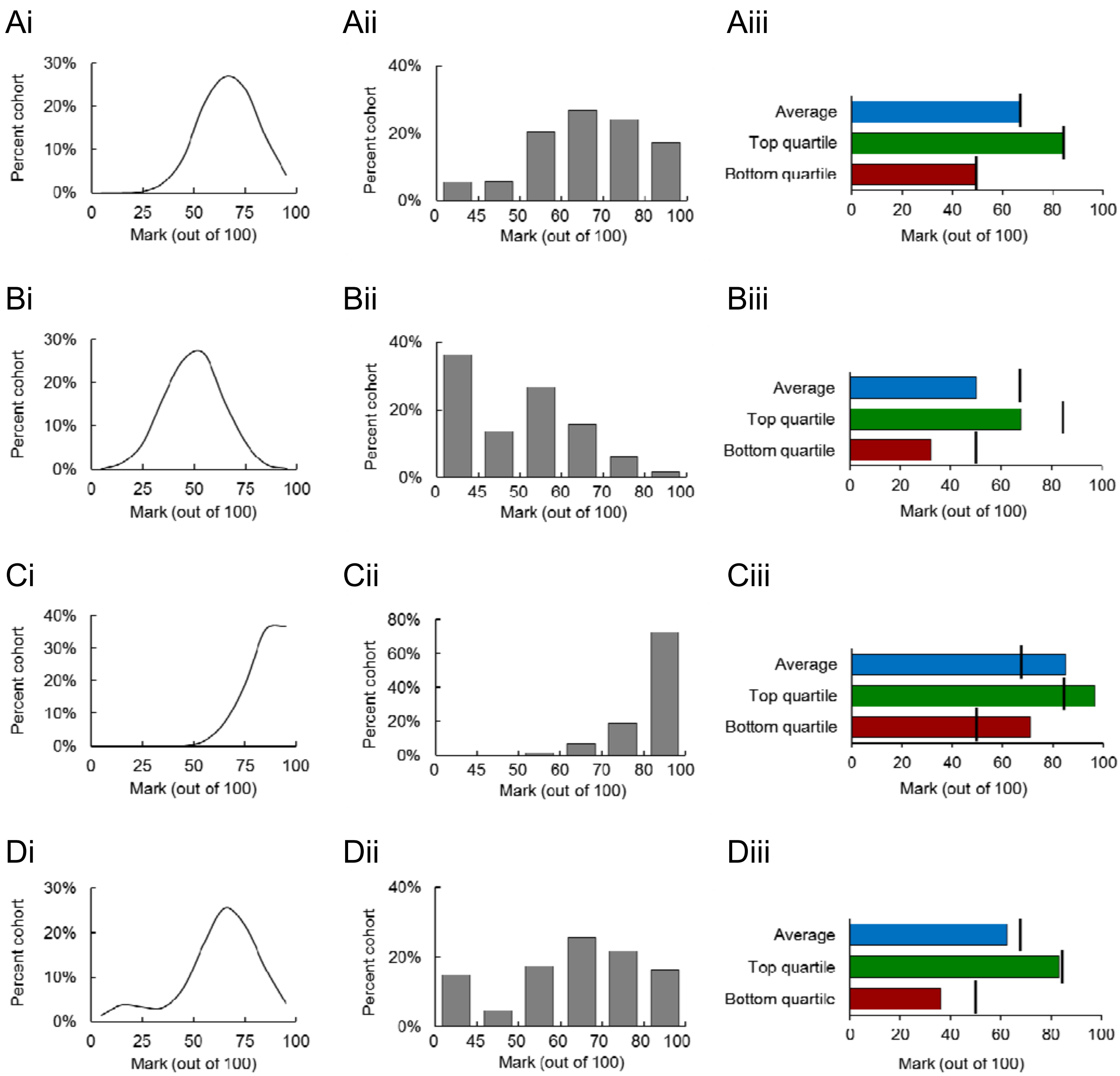
Grades are given as percentages. (A) Expanded GDD for Laboratory 1, year 1. Solid vertical lines show average grades for all laboratories for that component. To assist in comparisons, vertical lines show mean grades for the five laboratories for bottom quartile of students (red lines), average of all students (blue lines) and top quartile of students (green lines). (B) For the post-laboratory multiple-choice questions (MCQ), there were four MCQ. Expanded data shows the grades for each MCQ (MCQ1-4) visually and numerically (as a percentage) in the first year (year 1). (C) Grades for the post-laboratory MCQs in year 2 (see main text), following modification to laboratory content.

**REFERENCES**


1. G. Di Trapani, F. Clarke (2012) Biotechniques laboratory: An enabling course in the biological sciences. Biochem. Mol. Biol. Educ. 40, 29-36.
2. D. J. Adams (2015) Current Trends in Laboratory Class Teaching in University Bioscience Programmes. Bioscience Education 13, 1-14.
3. A. Boin, K. Elliott, H. Irving, V. Galea, E. Johnson. in R. Atkinson and C. McBeath, Ed. (2009) Same places, different spaces, 26th ASCILITE Conference Australasian Society for Computers in Learning in Tertiary Education (ascilite) Auckland, New Zealand pp. 81-90.
4. A. Bugarcic, K. Zimbardi, J. Macaranas, P. Thorn (2012) An inquiry-based practical for a large, foundation-level undergraduate laboratory that enhances student understanding of basic cellular concepts and scientific experimental design. Biochem. Mol. Biol. Educ. 40, 174-80.
5. K. Elliott, A. Boin, H. Irving, E. Johnson, V. Galea (2010) Teaching Scientific Inquiry Skills: A Handbook for Bioscience Educators in Australian Universities. Sydney, Australia: Australian Learning and Teaching Council.
6. P. Arthur, M. Ludwig, J. Castelli, P. Kirkwood, P. Attwood (2016) Prepare, Do, Review: A skills-based approach for laboratory practical classes in biochemistry and molecular biology. Biochem. Mol. Biol. Educ. 44, 276-87.
7. H. B. White, M. A. Benore, T. F. Sumter, B. D. Caldwell, E. Bell (2013) What skills should students of undergraduate biochemistry and molecular biology programs have upon graduation? Biochem. Mol. Biol. Educ. 41, 297-301.
8. B. Caldwell, C. Rohlman, M. Benore-Parsons (2004) A curriculum skills matrix for development and assessment of undergraduate biochemistry and molecular biology laboratory programs. Biochem. Mol. Biol. Educ. 32, 11-6.
9. A. Jenkins, M. Healey, R. Zetter (2007) Linking teaching and research in disciplines and departments. https://www.heacademy.ac.uk/knowledge-hub/linking-teaching-and-research-disciplines-and-departments.
10. C. A. Brewer, D. Smith, eds. (2011) Vision and change in undergraduate biology education, a call to action. American Association for the Advancement of Science: Washington, USA.
11. S. Brown (2004) Assessment for Learning. Learning and Teaching in Higher Education 1, 81-89.
12. S. Freeman, D. Haak, M. P. Wenderoth (2011) Increased course structure improves performance in introductory biology. CBE Life Sci Educ 10, 175-86.
13. R. Nagata (2004) The use of item analysis for improvement of biochemical teaching. Biochem. Mol. Biol. Educ. 32, 167-169.
14. T. R. Anderson (2007) Bridging the educational research-teaching practice gap: The power of assessment. Biochem. Mol. Biol. Educ. 35, 471-7.
15. S. L. Fowell, L. J. Southgate, J. G. Bligh (1999) Evaluating assessment: the missing link? Med Educ 33, 276-81.
16. D. R. Caprette, S. Armstrong, K. B. Beason (2005) Modular laboratory courses: An alternative to a traditional laboratory program. Biochem. Mol. Biol. Educ. 33, 351-5.
17. G. T. Crisp, E. J. Palmer (2007) Engaging academics with a simplified analysis of their multiple-choice question (MCQ) assessment results. Journal of University Teaching & Learning Practice 4, 88-106.
18. T. R. Anderson, J. M. Rogan (2010) Bridging the educational research-teaching practice gap: Tools for evaluating the quality of assessment instruments. Biochem Mol Biol Educ 38, 51-7.
19. G. Walsh (2010) The importance of and an approach to comprehensive reflective practice. Biochem. Mol. Biol. Educ. 38, 1-3.
20. M. Price (2005) Assessment standards: the role of communities of practice and the scholarship of assessment. Assessment & Evaluation in Higher Education 30, 215-230.
21. B. S. Malau-Aduli, D. Assenheimer, D. Choi-Lundberg, C. Zimitat (2014) Using computer-based technology to improve feedback to staff and students on MCQ assessments. Innovations in Education and Teaching International 51, 510-522.

22. R. D. Slater, M. Beal-Hodges, A. Reed (2014) Using excels pivot table function for visual data analysis of exam results: A supplemental procedure to classical test theory. Academy of Educational Leadership Journal 18, 221-229.
23. N. Iliinsky, J. Steele (2011) Designing Data Visualizations: Representing Informational Relationships. O'Reilly Media.
24. K. J. Schonborn, T. R. Anderson (2006) The importance of visual literacy in the education of biochemists. Biochem. Mol. Biol. Educ. 34, 94-102.
25. T. Ackerman (1996) Graphical Representation of Multidimensional Item Response Theory Analyses. Applied Psychological Measurement 20, 311-329.
26. M. McAlpine (2002) A summary of methods of item analysis. [23/10/2016]; http://caacentre.lboro.ac.uk/dldocs/Bp2final.pdf.
27. T. R. Anderson, J. M. Rogan (2011) Bridging the educational research-teaching practice gap. Curriculum development, Part 1: Components of the curriculum and influences on the process of curriculum design. Biochem. Mol. Biol. Educ. 39, 68-76.
28. T. R. Anderson (2007) Bridging the educational research-teaching practice gap: The importance of bridging the gap between science education research and its application in biochemistry teaching and learning: Barriers and strategies. Biochem. Mol. Biol. Educ. 35, 465-70.

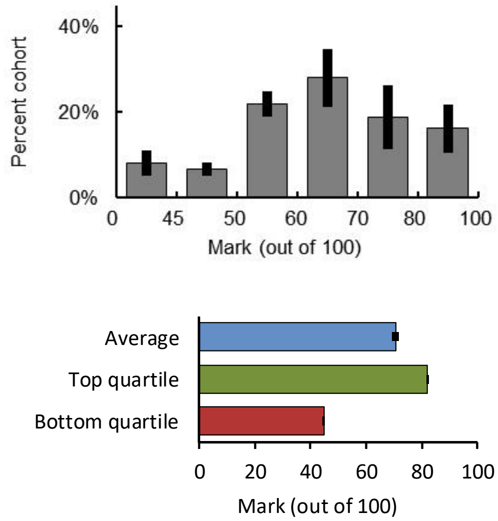
Figure 1


Ai
Aii
Aiii
Bi
Bii
Biii
Ci
Cii
Ciii
Di
Dii
Diii
Percent cohort
Mark (out of 100)
Average
Top quartile
Bottom quartile
0%
10%
20%
30%
40%
60%
80%
0
25
50
75
100
45
60
70
80
20
40

# Figure 2

A

B

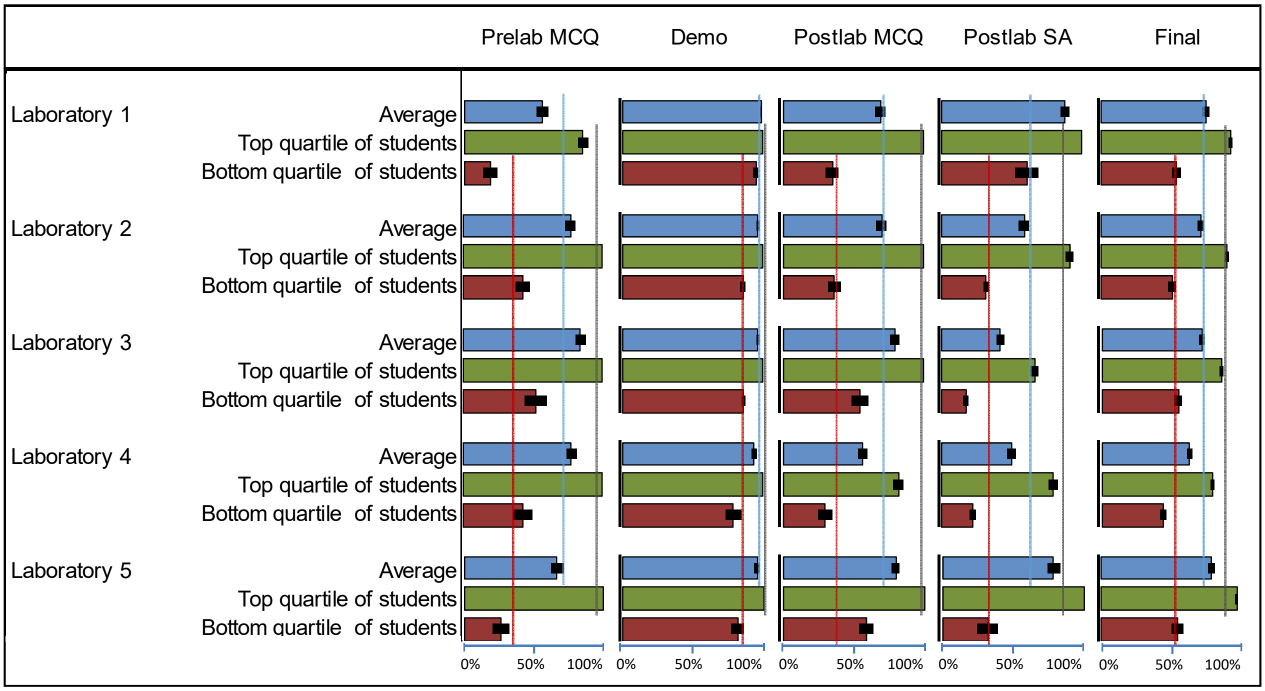



C

| | Mark (out of 100) | Confidence interval (95%) |
|---|---|---|
| Average | 70.5 | 1.3 |
| Top quartile | 82.1 | 0.3 |
| Bottom quartile | 44.8 | 0.5 |

# Figure 3

A

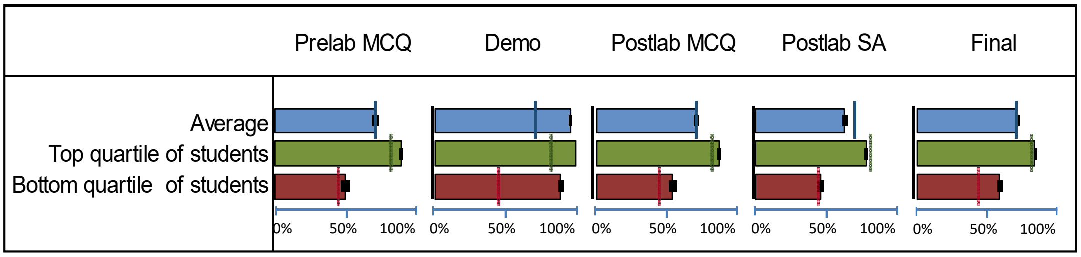



B

| | | Prelab MCQ | Demo | Postlab MCQ | Postlab SA | Final |
|---|---|---|---|---|---|---|
| Laboratory 1 | Average | 56 (4) | 99 (1) | 69 (4) | 88 (3) | 75 (2) |
| | Top quartile of students | 85 (4) | 100 (na) | 100 (na) | 100 (na) | 92 (1) |
| | Bottom quartile of students | 19 (5) | 95 (2) | 35 (5) | 61 (8) | 54 (3) |
| Laboratory 2 | Average | 77 (3) | 97 (1) | 70 (4) | 59 (4) | 71 (2) |
| | Top quartile of students | 100 (na) | 100 (na) | 100 (na) | 91 (3) | 90 (2) |
| | Bottom quartile of students | 43 (5) | 86 (2) | 37 (5) | 32 (2) | 51 (3) |
| Laboratory 3 | Average | 84 (4) | 97 (1) | 80 (3) | 42 (3) | 71 (2) |
| | Top quartile of students | 100 (na) | 100 (na) | 100 (na) | 66 (3) | 85 (1) |
| | Bottom quartile of students | 52 (8) | 86 (1) | 55 (6) | 17 (2) | 54 (3) |
| Laboratory 4 | Average | 78 (4) | 94 (2) | 57 (3) | 50 (3) | 63 (2) |
| | Top quartile of students | 100 (na) | 100 (na) | 82 (3) | 80 (3) | 79 (2) |
| | Bottom quartile of students | 43 (7) | 79 (5) | 30 (5) | 23 (2) | 44 (2) |
| Laboratory 5 | Average | 67 (4) | 95 (2) | 80 (3) | 78 (4) | 78 (2) |
| | Top quartile of students | 100 (na) | 100 (na) | 100 (na) | 100 (na) | 96 (1) |
| | Bottom quartile of students | 27 (6) | 81 (4) | 59 (5) | 32 (7) | 55 (4) |

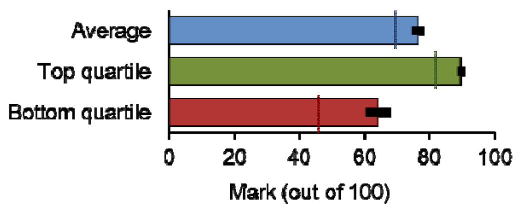
Figure 4

A




| | Prelab MCQ | Demo | Postlab MCQ | Postlab SA | Final |
|---|---|---|---|---|---|
| Average | 72 (2.3) | 96 (0.8) | 71 (2) | 64 (2) | 72 (1) |
| Top quartile of students | 90 (1) | 100 (na) | 87 (1) | 79 (1) | 84 (1) |
| Bottom quartile of students | 51 (3) | 89 (2) | 54 (3) | 46 (2) | 59 (2) |

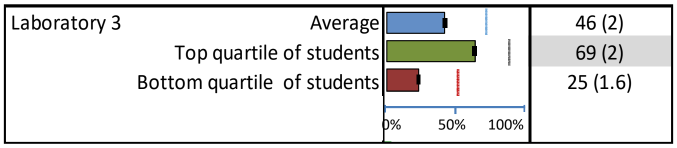
B



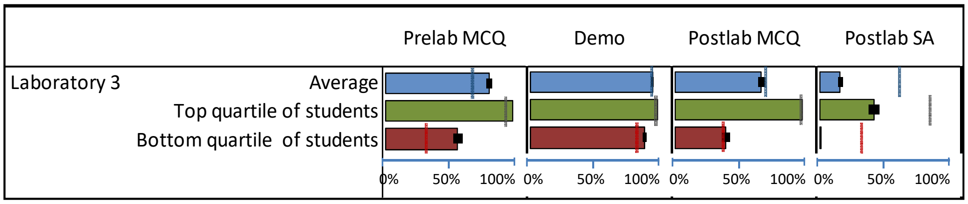

| | Demo |
|---|---|
| Average | 76 (1.5) |
| Top quartile | 90 (0.9) |
| Bottom quartile | 64 (2.3) |

# Figure 5

A

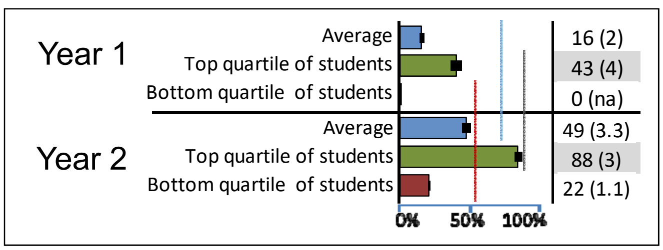



B



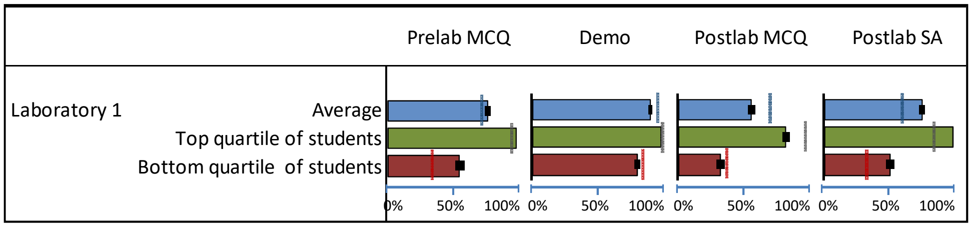

| | | Prelab MCQ | Demo | Postlab MCQ | Postlab SA |
|---|---|---|---|---|---|
| Laboratory 3 | Average | 82 (2.4) | 96 (0.7) | 68 (2.7) | 16 (2) |
| | Top quartile of students | 100 (na) | 100 (na) | 100 (na) | 42 (4) |
| | Bottom quartile of students | 57 (3.7) | 90 (1.5) | 40 (3.4) | 0 (na) |

C



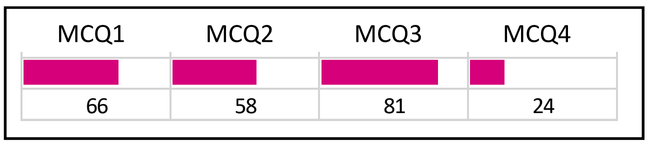
# Figure 6

A




| | | Prelab MCQ | Demo | Postlab MCQ | Postlab SA |
|---|---|---|---|---|---|
| Laboratory 1 | Average | 78 (2.3) | 92 (1) | 57 (2.6) | 76 (2.4) |
| | Top quartile of students | 100 (na) | 100 (na) | 84 (2.6) | 100 (na) |
| | Bottom quartile of students | 56 (3.7) | 82 (2) | 33 (2.9) | 51 (3.4) |

B




C