Changing Bimodal Grade Distributions – A Missed Opportunity?

Karamjeet K. Singh[1], Tara Allohverdi[1] & Steffen P. Graether[1]

[1] Department of Molecular and Cellular Biology, University of Guelph, Guelph, Ontario, Canada

Correspondence: Steffen P. Graether, Department of Molecular and Cellular Biology, University of Guelph, Guelph, Ontario, Canada. E-mail: graether@uoguelph.ca, ORCID 0000-0002-2356-3936

Received: March 10, 2022 Accepted: March 30, 2022 Online Published: April 1, 2022

doi:10.5430/ijhe.v11n5p70 URL: https://doi.org/10.5430/ijhe.v11n5p70

Abstract

Bimodal grade distributions indicate a gap in learning, where the highest quartile of students is skilful in the subject matter, but the lowest quartile is not retaining course material that demonstrates a good level of understanding. Students in the lower quartile do not necessarily have the same challenges as remedial students (i.e., those that do not meet the minimum course requirement) and should therefore be directed differently. We suggest that it is important to consider which elements of the course can be modified to reduce or eliminate bimodality. We provide here an approach to detect bimodality, explore the causes, and provide potential solutions that could be applied to any course. Our case study is on a third-year biochemistry course where several semesters showed a bimodal grade distribution. While student composition and timing of the course may have contributed to this result, underlying causes that can be controlled by the instructor include the lack of student engagement and academic motivation. Increasing the opportunities to earn marks and receive feedback, adding online components using social media, implementing seminars, and training teaching assistants to lead seminars can help reduce this problem.

Keywords: bimodal grades, assessment, attitudes, engagement, large classes

1. Introduction

Bimodal grade distributions in university courses occur when student performance falls into two clusters that consist of high and low grades. This contrasts with the classic bell curve, which has a unimodal distribution of student grades around a mean. Bimodal grade distributions can be a pedological issue; they indicate that there is a large gap in understanding between significant portions of the class. Additionally, students that fall into the lower grade portion are more likely to be disengaged, which could ultimately lead to attrition (Corney, 2009). Note that we distinguish here between the low-grade population in a bimodal distribution from remedial students, which can arise even with a normal distribution of grades. Remedial programs are structured to help students who have average or higher intellectual abilities but are not performing well because they may struggle with one specific area (e.g., reading, writing, or mathematics). In contrast, the lower grade population in a bimodal grade distribution may have an issue with understanding course content (i.e., intellectual ability).

Bimodal grades appear to be a universal problem and have been observed in many different geographical locations (Corney, 2009; Robins, 2010; Hook and Eckerdal, 2015; Gensheimer and Diebold, 2015); however, studies addressing this issue have mainly occurred in computer science courses (Robins, 2010; Hook and Eckerdal, 2015), with limited studies performed in other subjects. Multiple researchers have detected bimodality in several introductory computer science courses, but they present different solutions on how to address the problem. One study showed that final exam grades in an introductory programming class had a bimodal distribution (Hook and Eckerdal, 2015). The cause was explained as classes possessing two populations; students with an innate talent in programming versus those students who lack it. To identify if other factors could contribute to this distribution, the authors administered a questionnaire to the class to gather information on factors such as time spent on the course, attendance, personal views on a topic, and suggestions for course improvement. The results indicated that the major differences between the students that excelled or those that did not was that the students in the lower grade cohort spent less time working with a computer and found the topic difficult (Hook and Eckerdal, 2015). The authors proposed introducing lab-based supplemental instructions, where students who previously excelled in the course can mentor current students. This would not only allow the students to master the content, but also gives the newer students extra practice and time to ask questions on concepts that they find difficult, with an overall aim of improving student engagement (Hook and Eckerdal, 2015). Additionally, the authors commented on how classical examinations may not be the best way to grade programming skills, and that perhaps having more assessments in the form of assignments in which students had to solve programming tasks in a limited time could lead to a reduction in bimodality.

In another computer science programme study bimodal distributions were observed in introductory programming courses, where one proportion of students received a high mark and another received a low one (Corney, 2009). Researchers found this to be a very problematic issue because a consequence of the large number of poorly performing students resulted in a high level of attrition amongst first-year students (Corney, 2009). Previous attempts to ameliorate this bimodal distribution involved using different programming languages, paradigms, and teaching approaches, yet bimodality persisted (Corney, 2009). As a solution, the researchers suggest restructuring the program, with a major change being combining programming, database and web development into one first year course to allow students to gain an early understanding of the basic concepts, although at a more general level (Corney, 2009). The motivation was to increase student engagement early to establish the foundations to help with more difficult concepts later (Corney, 2009). This research was only proposed and the results of implementation have not been published.

The issue of bimodal grade distribution was also studied in a psychology course (Gensheimer and Diebold, 2015). The results showed that there were two nearly equal populations of low performers (D or F letter grade) and high performers (A or B letter grade). This observation motivated the researchers to explore potential causes for this bimodality, and the effect scoring poorly on the first exam had on the overall grade in the course. The results showed that academic major, attendance, and quiz performance distinguished the early low performers from the early high performers (Gensheimer and Diebold, 2015). The results also showed that about half of the low performers went on to receive a final grade of at least a C, with about one third of these students getting an A or B. The main factors that correlated with this improvement were attendance, subsequent performance on quizzes and exams, and doing extra credit. Overall, the authors concluded that engaging in basic academic behaviours that are expected in college and university, such as attending class, reading the assigned materials and completing homework, and making use of extra credit opportunities, led to student success. The authors also provided suggestions on a few areas to improve engagement to potentially reduce bimodality: mandatory attendance, low stake unannounced quizzes, extra-credit opportunities, early assessments to detect bimodality, and informing students of the research that links student engagement with success (Gensheimer and Diebold, 2015).

We have also observed the issue of bimodal grade distributions in a third-year undergraduate course (named here BIOC II) that builds on introductory biochemistry. We argue that the detection of bimodality can be easily determined in any science course using the method outlined here, and that there is a need for instructors to consider this as an issue which differs from failure and retention rates. Detecting bimodal grade distributions is an important initial tool to identify gaps in student learning, which can be the basis for further investigations to examine causes and search for solutions.

2. Method

The grades for various semesters from BIOC II, a follow-up to the introductory biochemistry course at the University of Guelph (Guelph, ON, Canada), were selected in this study because one of the authors (SPG) had observed a potential bimodal grade distribution. The University of Guelph is a medium-sized Canadian postsecondary institution, with a full-time undergraduate population of 23,926 in 2019/2020. Research Ethics Board (REB) permission (REB #17-12-004) was given to use grades without any association to student identity, which necessitated the binning of grades to a 5% range to remain in compliance. The semester of the grade could be identified; however, the year could not be disclosed.

To determine if grade distributions were unimodal, we developed a quantitative method using the Hartigan's dip test for unimodality (Hartigan and Hartigan, 1985). The data used was the final grade percentage grade in BIOC II from 10 different semesters. In this test, a p-value < 0.05 suggests that the null hypothesis (that the data are unimodally distributed) can be rejected, and the alternate hypothesis of a non-unimodal distribution can be accepted (Hartigan and Hartigan, 1985). We did not employ further tests to distinguish between bimodal and higher modalities, but simply assume that the data are at least bimodally distributed.

The steps we took to analyse the grade distribution for each semester in the software package R (R Core Team, 2013) are described below.

1. Download and install "quantmod", "gdata", and "diptest" packages in R.

2. Grade data for each semester in bins of 5%. While smaller bins can be used, we found that 5% bin widths allowed

for smoothing of the data. These data were used to create an .xlsx file for each semester.

3. Read the files for each semester in R as an .xlsx file.

4. Use the "count" function in R to make a count for each file. This will count the frequency of values in each bin.

5. A bar plot can be made using the "barplot" function in R to visualize the data.

6. The "dip.test" function was used to perform the Hartigan's diptest on each of the "count" files for the different semesters.

7. Analyse the result of the diptest by printing the results of the diptest using "print(*name of diptest file*). A p-value > 0.05 suggests a unimodal distribution, whereas a p-value < 0.05 suggests multi-modality (i.e., at least bimodal).

3. Results and Discussion

The grades for ten offerings of BIOC II over several contiguous years and semesters were analysed as outlined in the Methods. As suggested by casual inspection, the bar plots (Figure 1) showed that the grade distributions between different semesters could be bimodal. The fall semester grades were typically spread over a number of bins, while the winter semester grades seemed to show some right skewness (i.e., towards higher grades). It is difficult to objectively assess unimodality simply by eye, so the Hartigan's dip test was performed on all the available data to provide statistical support for the analysis. The results (Table 1) show bimodality or higher modality in several of the semesters. Three of the five fall semesters showed bimodal grade distributions, whereas all five winter offerings and the remaining two fall semesters were unimodal (Table 1). Note that one of the fall semester offerings, while above the cut-off, had a low p-value (0.088), showing that it was trending towards a non-unimodal distribution.

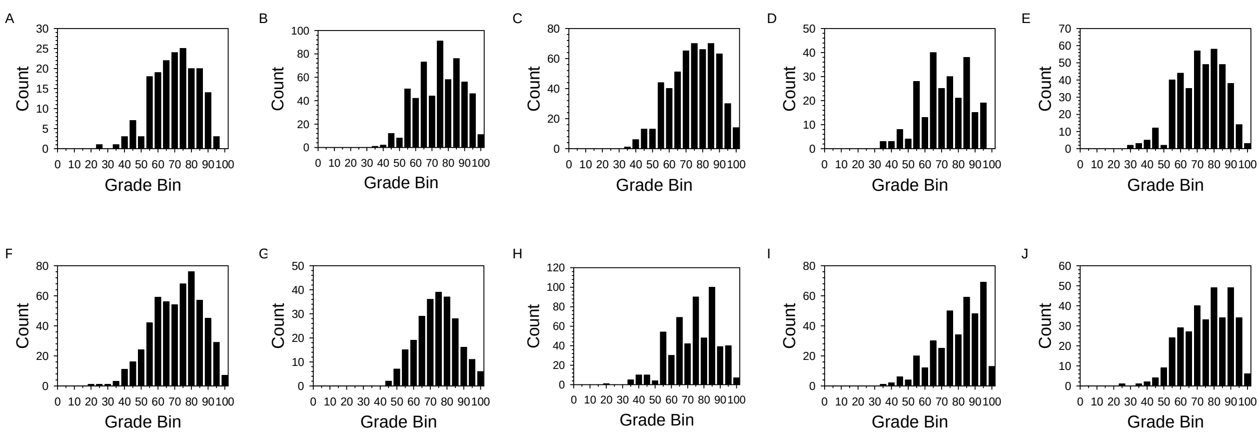

Figure 1. Distribution of Grades by Semester. Bar plot of the grades where the data are in 5% bins. The corresponding semester is shown in Table 1.

The difference between the fall and winter semesters is an interesting result, since there are differences between the two offerings. The most significant is that the student composition differs, with the winter semester being comprised of primarily Biochemistry, Microbiology, Molecular Biology & Genetics and Biomedical majors, who require this course, while the fall semester is primarily students retaking the course or other majors taking the course as a restricted elective. Another difference is that students that take the course in the winter semester have largely just taken the prerequisite, introductory biochemistry course in the previous semester, whereas almost all fall semester students would have last had the introductory course one or more years ago. Both factors are likely contributors to the different grade distributions that were observed; students in the winter semester are mainly in majors that are more likely to be interested in the content, or highly motivated to follow course content to be successful in applications to professional programs. Also, students that take BIOC II immediately after the introductory biochemistry are more likely to build off the recently learned content.

Table 1. Summary of Hartigan Dip Test

Semester	Class size	*p*-value for Dip's Test	Modality	Number of Peaks	Peak Bin Position
A (F)	180	0.003868	M	3	45,75,85
B (F)	570	0.02462	M	5	45,55,65,75,85
C (W)	546	0.1192	U	3	55,75,85
D (W)	247	0.8386	U	5	45,55,65,75,85
E (F)	411	0.02479	M	4	45,60,70,80
F (W)	550	0.3702	U	2	60,80
G (W)	245	0.4368	U	1	75
H (F)	549	0.2181	U	6	45,55,65,75,85,95
I (W)	373	0.8949	U	6	45,55,65,75,85,95
J (F)	342	0.08897	U	4	60,70,80,90

The grade distributions of BIOC II were tested as described in the Methods. Abbreviations: (F) Fall semester; (W) Winter semester; U Unimodal distribution; M Bimodal or higher distribution.

Efforts to help remedial students are a significant issue, but one which is well-studied and supported in many post-secondary institutions (e.g., Hollis, 2009; Cheng, 2011; Othman *et al.*, 2013; Hernandez *et al.*, 2019; Sanabria *et al.*, 2020). However, we feel that bimodal grade distributions represent a neglected opportunity to boost the level of student success. As a first step, we need to explore different approaches and determine which ones are the most likely to be successful. One possible cause is the inability of some students to be engaged with the content, which may be related to their academic motivation (Kahu and Nelson, 2018). Sometimes this issue can arise from there being only a few, heavily weighted assessments. Poor performance on a major assessment can make it difficult to recover, and heavily weighted assignments can also put students under intense pressure to perform well (Rau and Durand, 2000).

Having more assessments with alternative grading mechanisms is a possible solution, but it may not be that simple. Instead of using a system where marks are deducted, researchers implemented an approach in which grading was based on gainable experience points in a computer ethics course (Gehringer and Peddycord, 2013). This course format is different from the traditional format of a maximum grade of 100% and marks being lost for mistakes, but instead rewards students for the quantity and quality of completed tasks. Their results showed that despite implementing a very different marking scheme very different from what is standard, a bimodal distribution was still observed (Gehringer and Peddycord, 2013). Bimodality was thought to be caused by the long turnaround time on assessments, which affected students' academic motivation, and the lack of restrictions on how many assignments that could be completed, which led to many students putting in minimal effort and receiving a D or an F as their final grade. In contrast, a different population of students put in considerable effort and received an A, i.e., they were considered to have performed very well (Gehringer and Peddycord, 2013). Therefore, it might be important to find a middle ground, where there are multiple opportunities for students to be assessed and get feedback to promote engagement and motivation, so that any gaps in learning can be detected and addressed in a timely fashion, but at the same time ensure that there are not too many so that a student does not feel they are excessive and unnecessary.

In the BIOC II course, the assessments consisted mainly of a midterm (35%) and a final exam (55%) with four, low-stake quizzes (10%) as well. A better approach may be to add more assessments in the form of additional quizzes, graded assignments, mandatory attendance, graded clicker questions, and/or seminars. This would allow for more opportunities for the students to assess their learning over shorter periods with quick turnaround times, and understand what areas need improvement for them to be successful on the major assessments. The downside for students may be a heavier workload, which could lead to added stress and impact their performance in other courses as well.

Another approach to improve student engagement that has been explored is incorporating social media into a course. One study implemented online homework assignments to determine if this tool could increase student engagement (Junco *et al.*, 2011). The students were split into experimental (70 students) and control (55 students) groups, where the experimental group used the micro-blogging platform Twitter for various graded academic discussions. The authors were able to quantify engagement using a 19-item scale based on the Nation Survey of Student Engagement (Kuh *et al.,* 2001), which included demographic items, items inquiring about a student's technology use, and items that were included for forthcoming analyses (Junco *et al.,* 2011). The statistical analyses assessing the correlation between engagement and final grades showed that compared to the control group that did not use Twitter, students in the experimental group had a significantly greater increase in engagement (Engagement score difference of 5.12 ± 6.69 for experimental group and 2.29 ± 7.67 for control group when comparing the engagement scores before and after the experiment) and a higher final grade (Semester GPA of 2.79 ± 0.85 for experimental group and 2.28 ± 1.08 for control group; Junco *et al.,* 2011). This study provides experimental evidence that social media can be used to promote student engagement, but also emphasizes a more active and participatory role by faculty to increase the engagement.

Another approach is to consider adding a seminar component, since many science classes consist only of lectures and graded assignments. A study by Goodman and Pascarella details the effectiveness of seminars for first year students (Goodman and Pascarella, 2006). Although the focus was on students new to university because of low retention rates, it is worth applying these concepts to upper year courses as well, where retention can still be an issue. The main goal of including seminars was to increase academic performance and persistence through academic and social integration, with a long-term goal of increasing retention (Goodman and Pascarella, 2006). The evidence suggests that the seminars benefit a diverse population of students; males and females, minorities and majorities, various ages, different majors, on and off campus students, and low and high achievers (Goodman and Pascarella, 2006). In addition, engagement in seminars led to better studying habits and higher grades later in their program as well (Goodman and Pascarella, 2006). Therefore, it is worth considering adding a seminar component to improve student engagement and help with mastery of the course content, with the aim of developing and applying these skills throughout the program. Depending on the size of the course, the seminars may be led by the instructor or teaching assistants.

An area of concern when considering teaching assistants to lead a seminar is the variability in the students' experience. A study demonstrated that more experienced chemistry teaching assistants typically resulted in better undergraduate student academic performance (Kurdziel *et al.,* 2003). The study highlighted the importance of properly training teaching assistants through general university teaching training, but also specific course-based training (preparatory meetings, rehearsing student meetings, practicing marking), with the implementation of all these tools improving undergraduate student performance. We found a similar result when looking at student experiences in large classes compared to small classes, where in the end the instructors' engagement in a course was the most significant indicator of a positive experience (Barlow Cash *et al.,* 2017).

4. Conclusion

Bimodality in grade distributions can indicate issues with student comprehension and academic success, such that intervention should be considered. We have identified multimodality in BIOC II in several fall semester offerings, and believe student composition and timing of the course in different degrees are significant contributors to this result. A major underlying cause for bimodal grade distributions can be the lack of student engagement and academic motivation; we therefore provide several suggestions to improve student engagement and motivation: analysing mark breakdown and allowing ample opportunities for students to get feedback and earn marks, considering the addition of online components using social media, implementing seminars, and properly training teaching assistants to lead the seminars. Even though bimodality in grades have been rarely reported in sciences, we provide a method to detect it and suggestions to help minimize the phenomenon.

References

Barlow Cash, C., Letargo, J., Graether, S. P., & Jacobs, S. R. (2017). An Analysis of the Perceptions and Resources of Large University Classes. *CBE Life Sci Ed, 16*(2), 1-12. https://doi.org/10.1187/cbe.16-01-0004

Cheng, L-H. (2011). Enhancement of student learning performance using personalized diagnosis and remedial learning system. *Computers and education*, *56*(1), 289-299. https://doi.org/10.1016/j.compedu.2010.07.015

Corney, Malcolm W. (2009). Designing for engagement: building IT systems. In: ALTC First Year Experience Curriculum Design Symposium 2009, 5-6 February 2009, Queensland University of Technology, Brisbane.

Gehringer, E., & Pennycord, B. (2013). Grading by experience points: An example from computer ethics. *IEEE Frontiers in Education.* https://doi.org/10.1109/FIE.2013.6685097

Gensheimer, L. K., & Diebold, C. T. (2015). Bimodal inverted bell grade distribution: implications for instruction and student retention. *J. College Student Retention, 16*(3), 307-324. https://doi.org/10.2190%2FCS.16.3.a

Goodman, K., & Pascarella, E. T. (2006). First-year seminars increase persistence and retention: a summary of the evidence from How College Affects Students. *Academic OneFile, 8*(7).

Hartigan, J. A., & Hartigan, P.M. (1985). The Dip Test of Unimodality. *The Annals of Statistics, 13*(1), 70-84. https://doi.org/10.1214/aos/1176346577

Hernandez, J., Rodriguez, F., Hilliger, I., & Perez-Sanagustin, M. (2019). MOOCs as a Remedial Complement: Students' Adoption and Learning Outcomes. *IEEE transactions on learning technologies, 12*(1), 133-141. https://doi.org/10.1109/TLT.2018.2830373

Hollis, L. P. (2009). Academic Advising in the Wonderland of College for Developmental Students. *College student journal, 43*(1), 31.

Höök, L. J., & Eckerdal, A. (2015, April 9). On the Bimodality in an Introductory Programming Course: An Analysis of Student Performance Factors. 79-86. https://doi.org/10.1109/LaTiCE.2015.25

Junco, R., Heiberger, G., & Lokan, E. (2011). The effect of Twitter on college student engagement and grades. *Journal of Computer Assisted Learning, 27*(2), 119-132. https://doi.org/10.1111/j.1365-2729.2010.00387.x

Kahu, E. R., & Nelson, K. (2018). Student engagement in the educational interface: understanding the mechanisms of student success. *Higher Education Research & Development, 37*(1), 58-71. https://doi.org/10.1080/07294360.2017.1344197

Kuh, G. D. et al. (2001). National survey of student engagement: the college student report: NSSE technical and norms report. *Bloomington, IN: Indiana University Center for Postsecondary Research and Planning.*

Kurdziel, J. P., Turner, J. A., Luft, J. A., & Roehrig, G. H. (2003). Graduate Teaching Assistants and Inquiry-Based Instruction: Implications for Graduate Teaching Assistant Training. *Journal of Chemical Education, 80*(10), 1206. https://doi.org/10.1021/ed080p1206

Othman, A., Fadel, H. M., & Shuqair, K. M. (2013). Effectiveness of the Remedial Courses on Improving EFL/ESL Students' Performance at University Level in the Arab World. *International journal of higher education, 2*(3), 1-7. https://doi.org/10.1021/ed080p1206

R Core Team. (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rau, W., & Durand, A. (2000). The academic ethic and college grades: Does hard work help students to" make the grade"? *Sociology of Education*, *73*, 19-38. https://doi.org/10.2307/2673197

Robins, A. (2010). Learning edge momentum: a new account of outcomes in CS1. *Computer Science Education, 20*(10), 37-71. https://doi.org/10.1080/08993401003612167

Sanabria, T., Penner, A., & Domina, T. (2020) Failing at Remediation? College Remedial Course taking, Failure and Long-Term Student Outcomes. *Research in higher education, 61*(4), 459-484. https://doi.org/10.1007/s11162-020-09590-z