Form@re
Open Journal per la formazione in rete

esperienze / riflessioni

Exploring the possibilities of automated feedback for third level students

Esplorare le possibilità di feedback automatizzato per gli studenti universitari

Ian Clancy[a], Ann Marcus-Quinn[b]

[a] *University of Limerick, Ireland, Ian.Clancy@ul.ie*
[b] *University of Limerick, Ireland, Ann.Marcus.Quinn@ul.ie*

Abstract

There are a multitude of methods of providing third level students with feedback on their performance. Feedback can take many forms, from handwritten comments, to individual verbal feedback delivered face to face, and some educators are providing recorded audio feedback delivered using an institutional Virtual Learning Environment. One of the challenges at third level is providing substantive and constructive formative and/or terminal feedback in a timely fashion. Students are becoming more self-directed in terms of their learning and they are becoming more demanding regarding feedback. This paper discusses the use of a free open source solution (Auto Multiple Choice, AMC) to provide constructive feedback to students in an effective and timely manner. This method of automated assessment feedback has been used by faculty in Physics at the University of Limerick to reduce the turnaround time in making comments available to students since 2014. The experience of using AMC has also resulted in a number of unanticipated results around student motivation and prediction of exam performance which are also discussed.

Keywords: automated feedback; flipped classroom; student motivation; teaching and learning.

Sintesi

Esistono numerosi metodi per fornire feedback agli studenti universitari sulle loro prestazioni. Il feedback può assumere molte forme, dai commenti scritti a mano al feedback verbale individuale in presenza. Alcuni educatori forniscono feedback audio registrati usando un ambiente di apprendimento virtuale istituzionale. Una delle sfide universitarie consiste nel fornire un feedback formativo e/o finale in modo tempestivo. Gli studenti stanno diventando semore più "apprendisti" auto-diretti e stanno diventando sempre più esigenti in termini di feedback. Il contributo riflette sull'uso di un dispositivo gratuito open source (Auto Multiple Choice, AMC) per fornire feedback costruttivi agli studenti in modo efficace e tempestivo. Il metodo di feedback di valutazione automatizzato è stato utilizzato dalla facoltà di Fisica dell'Università di Limerick per ridurre i tempi di consegna dei commenti disponibili per gli studenti dal 2014. L' utilizzo dell'AMC ha inoltre consentito di realizzare una serie di risultati imprevisti sulla motivazione degli studenti e sulla previsione delle prestazioni degli esami, discussi nel paper.

Parole chiave: feedback automatizzato; flipped classroom; motivazione; apprendimento.

1. Introduction

Appropriate and timely assessment and feedback is at the heart of any good learning experience at third level. Technology, now more than ever, enables faculty to be creative in terms of their assessment models and their provision of feedback to students. In recent years there has been a noted uptake of innovative solutions which exploit the available technologies in order to improve the student experience. The flipped classroom, once deemed to be quite subversive has become an established teaching methodology (Bevilacqua & Campión, 2019; Gilboy, Heinerichs, & Pazzaglia, 2015; Zainuddin & Halili, 2016) and as new technologies become available the teaching and learning experience adapts and evolves. One such emergent trend is the use of automated feedback.

However, it is not simply a matter of uploading lecture materials to a Virtual Learning Environment (VLE) and grading them by automated multiple choice tests (Davidson, 2017). There are many advantages in using automated feedback solutions such as Auto Multiple Choice (AMC) (Clancy & Marcus-Quinn, 2015). A software tool such as AMC enables lecturers to individualise assessment options such as quizzes.

A major part of any learning or assessment activity is the degree to which students receive timely and effective feedback (Blayney & Freeman, 2004). Automatic feedback offers enormous potential to both educators and students (Debuse & Lawley, 2016; Jordan & Mitchell, 2009; Maguire, Maguire, & Kelly, 2017). Ala-Mutka (2005) lists the advantages of automation as the speed, availability, consistency and objectivity of assessment. However, using such automatic tools does not diminish the need for careful pedagogical design of the assignment. Automated feedback can be used as part of a wider strategy to encourage students to invest more of their time preparing for class and preparing for the brief assessments (Edgcomb, Vahid, Lysecky, & Lysecky, 2017). In a study undertaken by Wilson and Czik (2016) where they compared the effects of combining teacher feedback with automated feedback they found that teachers using automated feedback gave proportionately more higher-level feedback and that the combined feedback was associated with greater student motivation.

In spite of the benefits afforded by automated feedback there is a low uptake rate and this may be attributed to a heavy workload allowance and lack of training in the area. (Debuse & Lawley, 2016). Automated feedback is relatively new to some disciplines. In 2016 Sharples et al. published a report describing new forms of teaching, learning and assessment to meet the evolving learner expectations. They included automated feedback as one of ten innovations available but not yet so widely used so as to have had a notable influence on education.

One study carried out by Wieling and Hofman (2010) using regression analysis reported that automated feedback did not show a substantial effect on student performance. (Manoharan, 2019) also found that cheating in multiple-choice examinations can be reduced by personalising questions and answer options. Manoharan reported that their experience of using personalisation mitigated cheating, but also encouraged students to focus on concepts rather than mere answers.

This paper discusses the use of a free open source solution (AMC) to provide constructive feedback to students in an effective and timely manner. This method of automated assessment feedback has been used by faculty in Physics at the University of Limerick to reduce the turnaround time in making comments available to students since 2014. Student feedback has consistently welcomed the low stakes quiz format with rapid turn-around time.

1.1. Lecture and Laboratory sessions

The physics module is delivered to a combination of second year physics students and fourth year physics education students and is delivered over a twelve week period with each week comprising three lectures, one tutorial hour and a two-hour experimental laboratory session. The students are assessed by a combination of continuous assessment and summative written examination. The continuous assessment comprises assessment of experimental laboratory performance and in-class written quizzes. Automated quiz creation and feedback is implemented for these in-class written quizzes.

The advantages of implementing automated quiz creation and feedback are many:

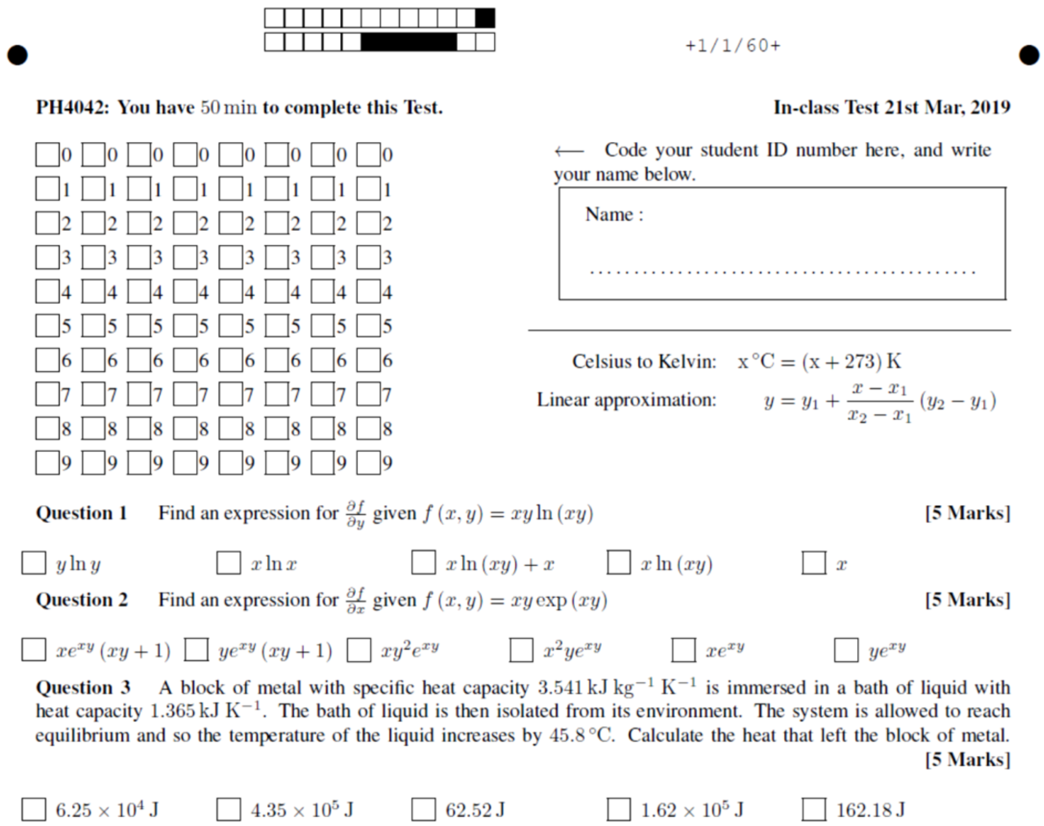
- the time to assess student responses is reduced considerably usually allowing feedback to be provided in the next session;
- all mark-up is delivered via the VLE and is typed enabling students to read and process individual feedback in private without confusion;
- as records are electronic, all records of original submissions could be retained to facilitate longitudinal study of future cohorts;
- the opportunity for students to copy from one-another is eliminated even in a crowded classroom setting;
- the difficulty of the quizzes motivates students to work systematically and continuously through the material and prepares them for the summative final examination;
- students can work collaboratively and engage in peer support by pooling their assessments and identifying shortcomings in their knowledge through repeated attempts at different problems.

2. Methodology

AMC is free software designed to create and manage multiple choice questionnaires. The AMC program was used to generate the set of questionnaires as a PDF document for printing. While AMC can be implemented on its own its real potential is achieved by combining it with LaTeX, an advanced typesetting platform common in physics and mathematics communities, and Python, a scripting programming language. The creation of these assessments and generation of automated feedback required the use of multiple free software packages. There is a steep learning curve in using these pieces of software individually and even more so when making them work in concert.

LaTeX is free software used for preparing documents and is particularly popular among academics in the fields of mathematics and physics. This software offers many advantages for typesetting large documents but particularly for formatting of complicated mathematical equations. The package pythontex allows the incorporation of Python programming code within LaTeX documents. This allows for the random generation of numbers for use in questions, print formatting of the numbers (significant figures used and scientific notation), and then the calculations based on these numbers to generate the correct answer and also incorporate common errors to the calculations for generating alternative incorrect answers. This approach can allow the examiner to identify errors that can be made in the calculation, coding them into the answers and then, based on the most common answers given by student, identifying the relative frequency of these common errors.

Such software has been used effectively to support the teaching and assessment of programming fundamentals (Ala-Mutka, 2005; Gotel & Scharff, 2007; Keuning, Jeuring, & Heeren, 2018) and mechanics (Suhonen & Tiili, 2014). Students had to complete between two and five assessments. Each of these assessments correspond to between two to five week blocks of lectures. The material associated with a given block tested concepts, methods, and theory but emphasized calculation-based questions. Assessments typically involved four to six calculation-based questions with one theory-based derivation. Students were then given thirty to fifty minutes to complete the quiz, depending on what was to be assessed. A partial example of the output of an AMC document is given in Figure 1.



+1/1/60+

PH4042: You have 50 min to complete this Test. **In-class Test 21st Mar, 2019**

☐0	☐0	☐0	☐0	☐0	☐0	☐0	☐0
☐1	☐1	☐1	☐1	☐1	☐1	☐1	☐1
☐2	☐2	☐2	☐2	☐2	☐2	☐2	☐2
☐3	☐3	☐3	☐3	☐3	☐3	☐3	☐3
☐4	☐4	☐4	☐4	☐4	☐4	☐4	☐4
☐5	☐5	☐5	☐5	☐5	☐5	☐5	☐5
☐6	☐6	☐6	☐6	☐6	☐6	☐6	☐6
☐7	☐7	☐7	☐7	☐7	☐7	☐7	☐7
☐8	☐8	☐8	☐8	☐8	☐8	☐8	☐8
☐9	☐9	☐9	☐9	☐9	☐9	☐9	☐9

⟵ Code your student ID number here, and write your name below.

Name : ....

Celsius to Kelvin: $x\,^{\circ}\mathrm{C} = (x + 273)\,\mathrm{K}$

Linear approximation: $y = y_1 + \frac{x - x_1}{x_2 - x_1}(y_2 - y_1)$

Question 1 Find an expression for $\frac{\partial f}{\partial y}$ given $f(x, y) = xy \ln(xy)$ **[5 Marks]**

☐ $y \ln y$ ☐ $x \ln x$ ☐ $x \ln(xy) + x$ ☐ $x \ln(xy)$ ☐ x

Question 2 Find an expression for $\frac{\partial f}{\partial x}$ given $f(x, y) = xy \exp(xy)$ **[5 Marks]**

☐ $xe^{xy}(xy + 1)$ ☐ $ye^{xy}(xy + 1)$ ☐ xy^2e^{xy} ☐ x^2ye^{xy} ☐ xe^{xy} ☐ ye^{xy}

Question 3 A block of metal with specific heat capacity $3.541\,\mathrm{kJ\,kg^{-1}\,K^{-1}}$ is immersed in a bath of liquid with heat capacity $1.365\,\mathrm{kJ\,K^{-1}}$. The bath of liquid is then isolated from its environment. The system is allowed to reach equilibrium and so the temperature of the liquid increases by $45.8\,^{\circ}\mathrm{C}$. Calculate the heat that left the block of metal. **[5 Marks]**

☐ $6.25 \times 10^4\,\mathrm{J}$ ☐ $4.35 \times 10^5\,\mathrm{J}$ ☐ $62.52\,\mathrm{J}$ ☐ $1.62 \times 10^5\,\mathrm{J}$ ☐ $162.18\,\mathrm{J}$

Figure 1. Example output of compiled AMC document using LaTeX. The square array allows students to enter their ID.

Calculation-based questions were individualized to students by randomizing the numbers provided in the questions leading to different correct answers for different students. The ordering of different answers was also randomized so correct answers would not appear in the same place on the physical page. These questions were in the format of multiple-choice questions, with alternative incorrect answers corresponding to distinct errors that can be made during the calculation. Care needed to be taken in the design of the questions that the calculation errors did not accidentally lead to the correct answer.

2.1. Automated marking from completed answers sheets scans

After the quiz, scanned completed answer-sheets were marked automatically by AMC, using Optical Mark Recognition (OMR). In this experience the lecturer had access to a scanner and scanned each of the answer sheets using an automatic feed. However, it is possible to correct the quiz without access to a scanner as manual computer-aided data capture is available. Where this is the preferred choice the lecturer must click on boxes that are ticked on the completed sheets. This may be a better option for some where automated marking is not efficient and indeed was required for correction of theory-based questions that cannot be formed into a multiple choice format. AMC allows fully customizable scoring strategies to be defined. In this case study negative marking was not used and only the correct answer was awarded marks for calculation based questions. Partial marks were awarded for each milestone of a theory-based question. Theory based questions required manual marking but this was performed through AMC, which streamlined the process. Typically students completed the quiz on a Friday and received feedback on their own performance and the overall class performance the next day. Figure 2 outlines the process flow used.

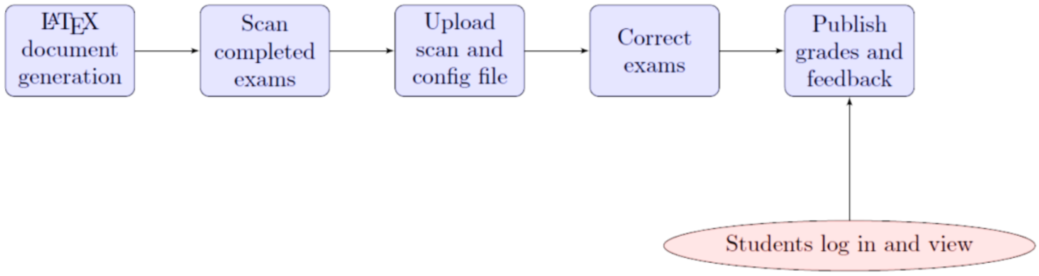


Figure 2. Hendricks' User functional model image (2012) outlines the flow for all stages of the automated feedback process.

3. Results and Discussion

The implementation of these individualised assessments and automated feedback has been run over six academic years denoted here as groups A to F.

Figure 3 shows the correlation plot between the final exam mark out of 100% and the overall quiz mark out of 100% for individual students for three of these groups. The scatter is consistent over the years with the Pearson correlation coefficient for each group varying within 95% confidence levels over the six years studied (

Figure 4). The correlation coefficient of all data is 0.634 and so does not suggest a strongly predictive relationship. However, the number of students who achieved a higher mark in the final exam compared to the quizzes was 54.7%. This suggests the quizzes motivated some students to improve their performance. This is supported by the distribution of the residuals, i.e. the difference between the overall quiz mark and the final exam mark. Figure 5 shows the distribution of residuals having a Gaussian distribution with a mean of -1.74 suggesting the average student got 1.74 percentage points more in the final exam than the quizzes. Again the distribution of residuals is quite broad and so the predictive power of this measure is limited.

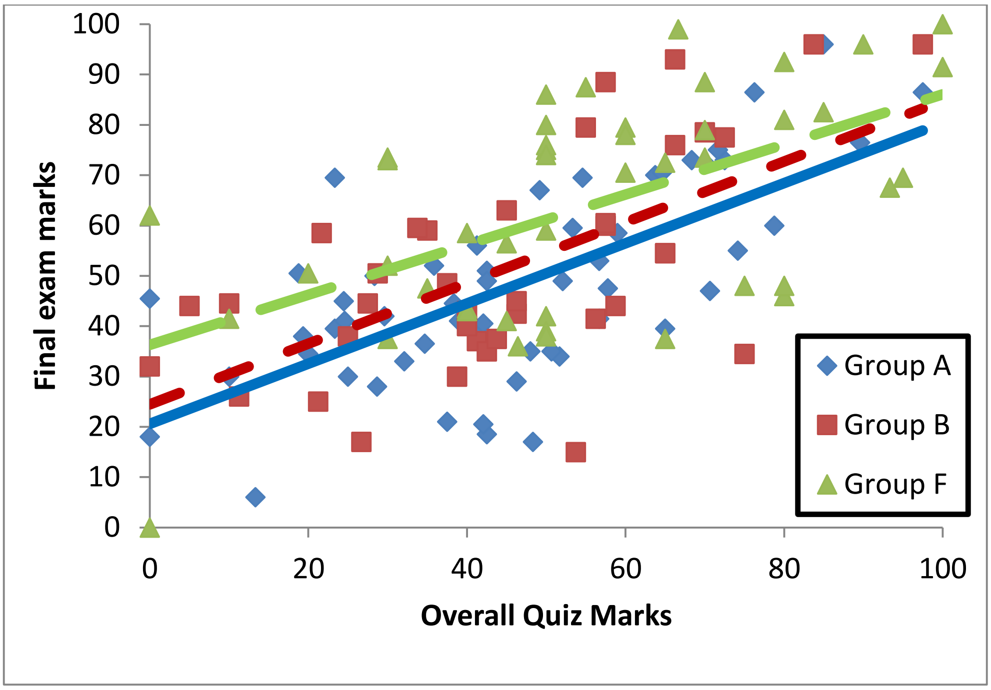


Figure 3. This figure shows the correlation plot between final exam marks and the overall quiz marks for groups A, B and F. The lines represent linear regression fits to the data for group A (solid line) group B (short dashed line) and group F (long dashed line).

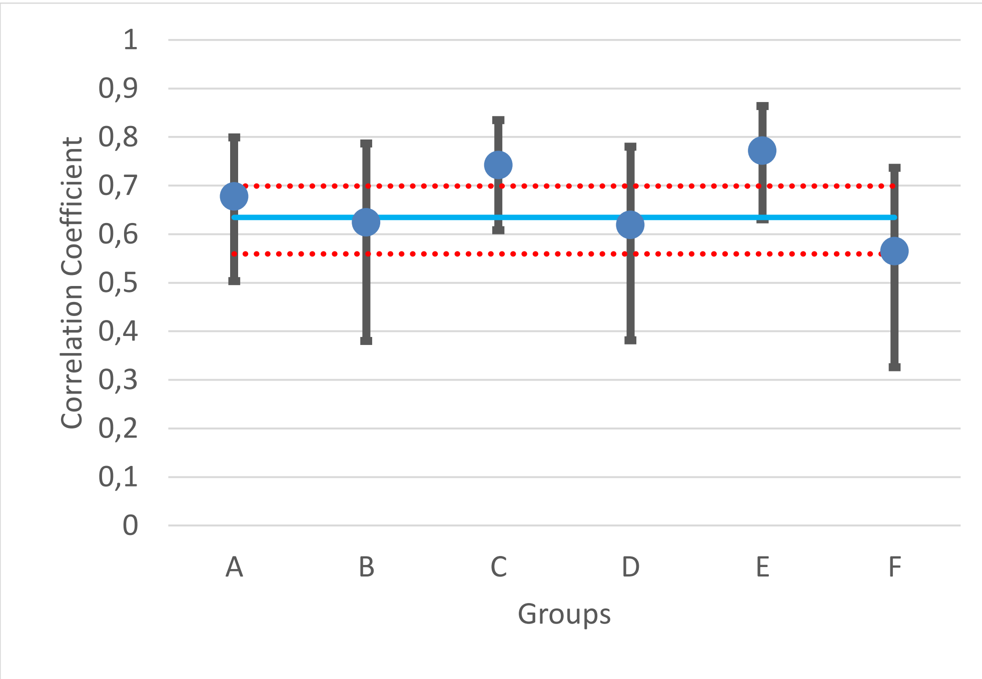


Figure 4. This figure shows the Pearson correlation coefficients and 95% control limits indicated by error bars for each group. The solid blue line represents the correlation coefficient for all data and the dashed red lines represent 95% upper and lower control limits.

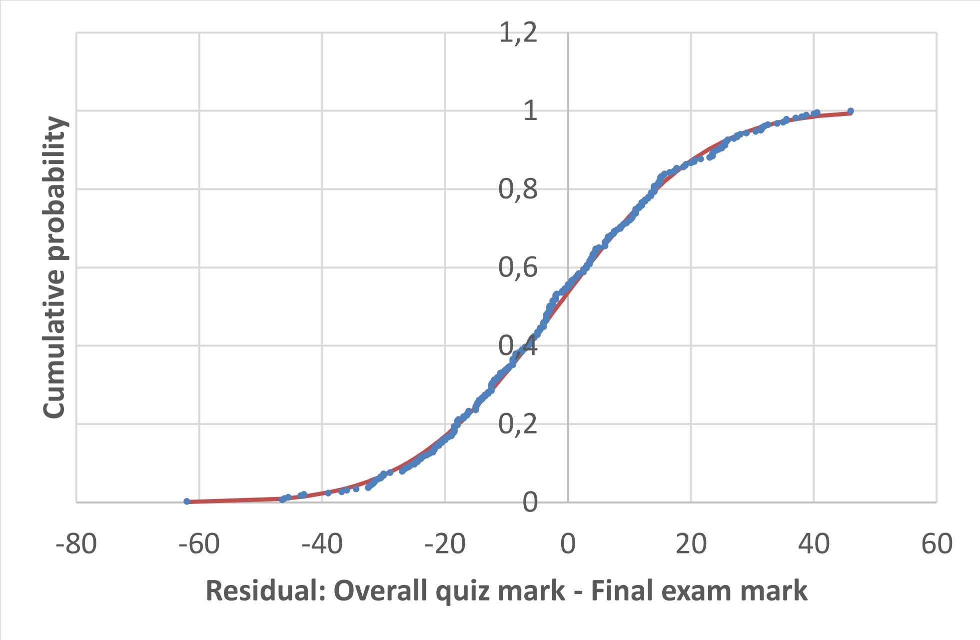


Figure 5. This figure presents the cumulative Gaussian distribution with mean -1.74 and standard deviation 27 (solid red curve) of the difference between the overall quiz percentage mark and corresponding final exam percentage mark. Each data point represents an individual student.

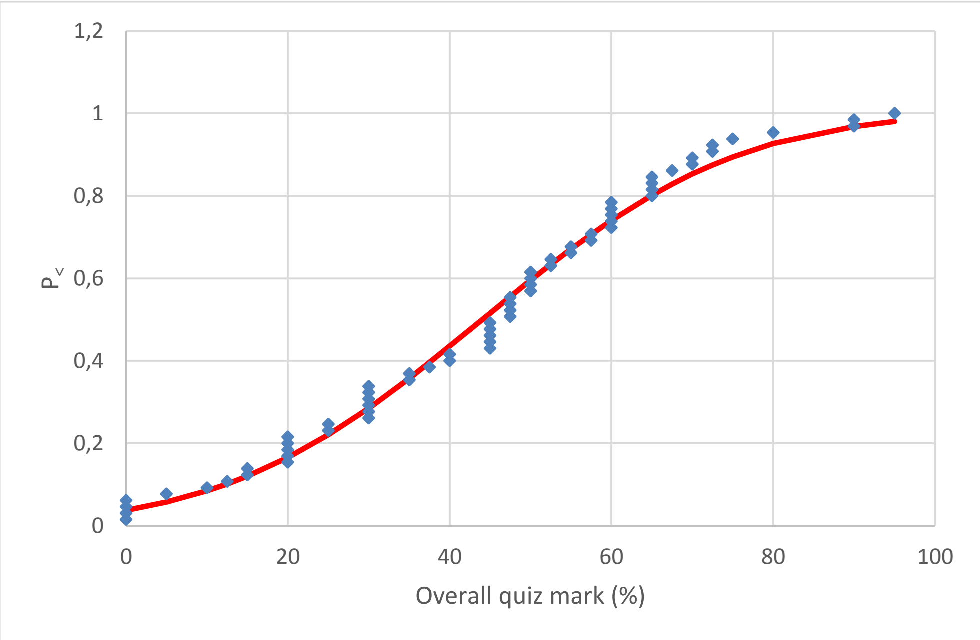


Figure 6. An example of one aspect of automated feedback provided to students showing the cumulative distribution of marks achieved.

An example of automated feedback provided to students is given in Figure 6 that represents the cumulative distribution of the quiz marks, *i.e.* for each student the fraction of the class that gets an equal or lower mark than the individual student. Therefore, the top student will have $P_< = 1$ while the median student will have $P_< = 0.5$. Each datapoint represents an individual student. The curve represents a cumulative distribution function of a Gaussian distribution (assumes normal distribution of marks) with mean 44% and standard deviation of 35%-points. This also allows the indication of mean and median class mark that students can compare with their own performance.

3.1. Challenges to implementation

There are a number of impediments to the incorporation of automatic creation of assessment and feedback. First among these is the steep learning curve in the proficient usage of LaTeX. In addition, Python is a programming language and so requires previous programming experience for a short lead-in time to use it efficiently with LaTeX.

The design and programming of questions themselves can take some time. First careful choices must be made about the range of random numbers assigned to problems as a bad choice could render the solutions nonsensical or indeed impossible. There is also the unintended effect of the incorrect solution algorithm giving the correct answer through a poor choice of numbers. This can lead to more than one multiple choice answer being identical.

The method of students entering their ID numbers can be confusing for some students and so this may require manual intervention to associate solutions to the correct student. The OMR technology requires a clear border to interpret the solutions so students doodling can interfere with this automated process. A forceful reminder to students before the start of the quiz can alleviate this problem.

Students should also be informed to fill in the boxes completely with a dark colour. Light pencil marks have too low a contrast for the software to reliably identify the student response. This could again require manual intervention to identify the answers chosen by the student.

It is important to have appropriate scanning facilities to reliably scan the student answers at a sufficient quality. An automated feed to scan to a single document in PDF is invaluable. The process of recognition and mark-up of these quizzes generates quite a lot of data with one quiz to about 40 students creating about 2 GB of data.

The return of marked-up quizzes was accomplished using the local VLE, however for large groups this may require a bespoke method of uploading the files for each student. In this instance a python script was created and used to distribute the files in the VLE’s preferred structure for bulk upload.

4. Conclusion

This paper discusses the potential and affordances offered by a free open source solution (AMC) to provide automated feedback to undergraduate physics students.

This experience in line with best international practice found that combining teacher and automated feedback can benefit both faculty and students. Using automated feedback enables teachers to provide feedback in a more timely and efficient manner at critical points during the teaching term.

The final results suggested a slightly better performance by students on average. This may suggest that the continuous assessment aspect of the automated feedback encouraged and motivated students to achieve a higher grade in the final exam. Currently automated feedback is under represented in the literature and it is hoped that in providing a report of a successful experience for both students and faculty that more academics may choose to implement some aspect of automated feedback into their teaching practice.

Reference list

Ala-Mutka, K. M. (2005). A survey of automated assessment approaches for programming assignments. *Computer science education*, *15*(2), 83–102.

Bevilacqua, A., & Campión, R. S. (2019). The Flipped Learning approach: quantitative research on the perception of Italian teachers. *Form@re - Open Journal per la formazione in rete, 19*(2), 405–422.

Blayney, P., & Freeman, M. (2004). Automated formative feedback and summative assessment using individualised spreadsheet assignments. *Australasian Journal of Educational Technology*, *20*(2), 209–231.

Carruthers, C., McCarron, B., Bolan, P., Devine, A., McMahon-Beattie, U., & Burns, A. (2015). 'I like the sound of that' – an evaluation of providing audio feedback via the virtual learning environment for summative assessment. *Assessment & Evaluation in Higher Education, 40*(3), 352–370.

Clancy, I., & Marcus-Quinn, A. (2015). Exploring the possibilities of automated feedback for undergraduate students (pp. 7264-7269). *EDULEARN15 Proceedings*, Barcelona, Spain. https://library.iated.org/view/CLANCY2015EXP (ver. 10.12.2019).

Davidson, C. N. (2017). *The new education: How to revolutionize the university to prepare students for a world in flux*. New York, NY: Hachette.

Debuse, J. C., & Lawley, M. (2016). Benefits and drawbacks of computer- based assessment and feedback systems: Student and educator perspectives. *British Journal of Educational Technology*, *47*(2), 294–301.

Dixon, S. (2015). The pastoral potential of audio feedback: A review of the literature. *Pastoral care in Education*, *33*(2), 96–104.

Edgcomb, A., Vahid, F., Lysecky, R., & Lysecky, S. (2017). Getting students to earnestly do reading, studying, and homework in an introductory programming class (pp. 171-176). *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education,* Seattle, Washington, USA.

Fawcett, H., & Oldfield, J. (2016). Investigating expectations and experiences of audio and written assignment feedback in first-year undergraduate students. *Teaching in Higher Education*, *21*(1), 79–93.

Gilboy, M. B., Heinerichs, S., & Pazzaglia, G. (2015). Enhancing student engagement using the flipped classroom. *Journal of Nutrition Education and Behavior*, *47*(1), 109–114.

Gotel, O., & Scharff, C. (2007). Adapting an open-source web-based assessment system for the automated assessment of programming problems. *IASTED Web-based Education Conference*, Chamonix, France.

Hendriks, R., (2012). Automatic exam correction. *UVA Universiteit van Amsterdam*. https://esc.fnwi.uva.nl/thesis/centraal/files/f438164865.pdf (ver. 22.11.2019).

Jordan, S., & Mitchell, T. (2009). eAssessment for learning? The potential of shortanswer free text questions with tailored feedback. *British Journal of Educational Technology, 40*(2), 371–385.

Keuning, H., Jeuring, J., & Heeren, B. (2018). A systematic literature review of automated feedback generation for programming exercises. *ACM Transactions on Computing Education (TOCE)*, *19*(1), 3.

Maguire, P., Maguire, R., & Kelly, R. (2017). Using automatic machine assessment to teach computer programming. *Computer Science Education, 27*(3-4), 197–214.

Manoharan, S. (2019). Cheat-resistant multiple-choice examinations using personalization. *Computers & Education*, *130*, 139–151.

Sharples, M., de Roock, R., Ferguson, R., Gaved, M., Herodotou, C., Koh, E., … & Wong, L. H. (2016). *Innovating Pedagogy 2016: Open University Innovation Report 5*. Milton Keynes: The Open University.

Suhonen, S., & Tiili, J. (2014). Combining good practices in fully online learning environment–introductory physics course. *Proceedings of SEFI2014 42nd Annual Conference,* Birmingham, UK.

Wieling, M. B., & Hofman, W. H. A. (2010). The impact of online video lecture recordings and automated feedback on student performance. *Computers & Education*, *54*(4), 992–998.

Wilson, J., & Czik, A. (2016). Automated essay evaluation software in English Language Arts classrooms: Effects on teacher feedback, student motivation, and writing quality. *Computers & Education*, *100*, 94–109.

Zainuddin, Z., & Halili, S. H. (2016). Flipped classroom research and trends from different fields of study. *The International Review of Research in Open and Distributed Learning*, *17*(3), 313–340.