STUDENT ACCESS TO AUTOMATED MARKING TOOLS AS A WAY OF IMPROVING THEIR COMPREHENSION

David Cutting[1], Andrew McDowell[1], Neil Anderson[1], Angela Allen, & Matthew Collins[1]
[1]Dr.,
School of Electronics, Electrical Engineering and Computer Science, Queen's University Belfast
(United Kingdom)

Abstract

Feedback, on both a summative and formative basis, is a critical component of modern teaching. At the same time as institutional expectations on quality and volume of feedback grow, learner expectations of turnaround times also become increasingly hard to meet, especially in large cohorts. One approach is to provide automatic feedback which can be applied on a formative basis or to support summative marking. To support learner experience on one large module a version of the automated summative marking system was provided on a formative basis to learners throughout their project offering a representative subset of all the marking criteria used and detailed feedback on errors found in submission. This tool has been offered over two student cohort years, 2018 and 2019, and been widely used by the learners. In 2019 there were a total of 13,141 unique submissions to the tool from at least 312 individuals (of 362 enrolled). Feedback from both learners and academic staff has been very positive, helping both pro-active problem solving and domain knowledge in the learners while reducing workload for staff, showing the approach has been highly effective in this domain.

Keywords: Feedback, automation, problem solving.

1. Introduction

Feedback is an essential part of the student journey of learning and progression providing the opportunity to see where and how they can improve their understanding (Leckey & Neill, 2001). Such feedback can be either summative or formative in nature with the vital component being the ability to "feed forward" with feedback being used to inform the learner in their next steps not simply if they were right or wrong (Quinton & Smallbone, 2010). Providing, timely, consistent, and clear feedback to students however remains a challenge, especially when dealing with large class sizes and tight deadlines between assessments. In such large cohorts there will naturally also be a wide range in learner preferences for the type and volume of feedback provided (Rowe & Wood, 2008) in addition to a continued pressure for fast turnaround times (Milanova et al., 2018).

One approach that may be employed in some domains is the use of automated tools, whereby assessment of submissions is performed by software, usually by checking submissions against a sample solution (Joy & Luck, 1998; Blumenstein, Green, Nguyen, and Muthukkumarasamy, 2004). Automation itself is not uncommonly used in computer science and tools are often applied by the marking staff for summative assessment after submission, a useful tool to reduce turnaround time and increase consistency but not in itself a facility directly used by the learner. Automated tools used directly by learners are normally formative in nature offering immediate feedback and forming part of the education process itself (Gerdes, Heeren, Jeuring, and van Binsbergen, 2017; Heckman & King, 2018).

With summative marking tools in place an option is available to provide access in some form to the automated tools to learners before submission, providing instantaneous formative feedback which is directly aligned to some or all of the same criteria with which their summative marks will be given. Providing such access directly to learners would also allow each individual to make their own use of the tools according to their preferences (Rowe & Wood, 2008).

To this end in 2018 a first-year module project was set to c. 400 learners to implement a database. While the intention had always been to automatically generate the summative marking for this work using tools (a bot) by comparing weighted requirements against a specimen that met all the requirements, an opportunity was identified to provide learners with early access to the bot. This would allow learners to

upload their proposed solutions and run these against a subset of the real marking tests (not all tests were included but there was good coverage of the types and areas to be tested) providing detailed output and a score of marks awarded against the theoretical maximum for their submission. In this way any learner could take advantage of the facility and have portions of their work examined, graded, and returned with feedback showing exactly which areas were lacking and needed further attention. The feedback would be sufficient to allow common areas to be inferred and corrected even in elements for which feedback was not directly provided. The intention was to reduce the load on staff while allowing learners to identify weaknesses and self-address those themes or concepts as required, resubmitting work for further feedback once changes had been made. Having been implemented and used in 2018 the same tools and techniques were re-applied to the next years cohort of learners in 2019.

2. Design and implementation of automated assessment tools

There were three main components to the databases submission comprised of SQL statements; the creation of a schema of 12 tables from a provided data dictionary, the insertion of given sample data into each table, and a set of ten queries of varying complexity to exploit the database. To facilitate automatic assessment a full sample solution was created and verified by staff against the requirements. An automated analysis was then performed to generate a set of meta files (in JSON format) showing all of the properties within the sample solution including names and types. Further tests were defined to check specific functionality which wasn't immediate obvious from the meta files, for example were the correct constraints in place to stop data being illegally inserted into the database. Finally each of these elements, several hundred in total, was assigned a relative score weighting. Minor items had small scores so if missed would have little impact whereas more major or key elements had much higher weighting. In total the available marks exceeded 2,000 across all the checked areas. A software system, AME (Automated Marking Engine), was them implemented to compare submissions against the sample solution using the defined scoring. For each submission a temporary protected database was created and used to run the submitted SQL before being checked.

With these definitions in place and an offline AME process complete to take a submission in three parts and score it, focus turned as to how the service could and should be exposed to learners to use. Having always desired to not make all facets of the assessment available each of the items tested had a flag to indicate if it should be made "public" (available through an external interface or not). Each test was considered and a decision made if it should be available to learners. The intention was to provide a meaningful subset of tests of different types to allow all common errors and checks to be at least inferred. For example not all tables were fully tested but some were showing that all fields including types and indexes were to be checked. All the ten output queries were included in the learner-facing test set as they varied significantly from each other.

One area of discussion at this point was the possibility for "follow-on failures" in learner submissions. An error in the schema for example could stop data being loaded and break the queries extracting the correct data. An error in one part could therefore cause perfectly correct code to later fail. To avoid this a decision was made to test each submission section against both the previous elements of the submission and again using the previous elements from the sample solution – using the score from whichever scored highest.

Once the determinations on learner-facing tests had been made a web interface, the Automated Marking Engine for You (AMEY), was implemented and hosted on the school network. This would provide a simple facility for learners to upload their submissions, in the same format as required for final project submission, and generate feedback. Each of the three stages would be progressive in that it would be possible for example to just test the schema component and see a score against the maximum for just that component before adding more elements. In the AMEY presentation just the submitted work was tested, there was no attempt to test later elements with the earlier elements being provided from the sample solution as this would potentially expose the sample solution SQL. Each test element was highlighted in the output and colour coded (green when all elements passed, yellow when some passed, and red when none passed).

To facilitate questions being asked to staff, each run through AMEY had an associated unique alphanumeric Session ID (SID) which was displayed. Learners could save this themselves to refer back to previous runs but could also provide this to staff when asking questions. Staff were then able to see in detail the complete student submission and marking output making it much more straightforward to identify issues and provide assistance to learners.

3. Learner use of automated tools

AMEY was provided to learners in both the 2018 and 2019 cohorts. In both years throughout the duration of the project significant use of the system was made, unsurprisingly peaking near to the date of submission. In total there were 6,231 submission in 2018 and 13,141 in 2019. Submission statistics for AMEY in 2018 are shown in Figure 1 and the same data for 2019 in Figure 2. In 2018 AMEY was only available on the campus network or through a remote virtual desktop environment, something with was identified as a barrier in feedback and addressed through a firewall change in 2019 which may explain the greater volume of submissions.

In 2018 there was no facility to capture which learners were using AMEY but in 2019 an extra field for student number was introduced. To avoid barriers to adoption this field wasn't validated or required and so the data is lacking for a number of submissions but from this it is clear that at least 312 individual learners (of 362 enrolled) made at least one submission to AMEY. Some individuals made high numbers of submissions (one made 483) and seven learners made just a single submission each. The mean was 37.5 submissions per valid student number entered with a standard deviation of 45.8.

Figure 1. AMEY statistics for 2018 cohort showing number of submissions along with both the average score and average maximum achievable scores (of elements submitted).
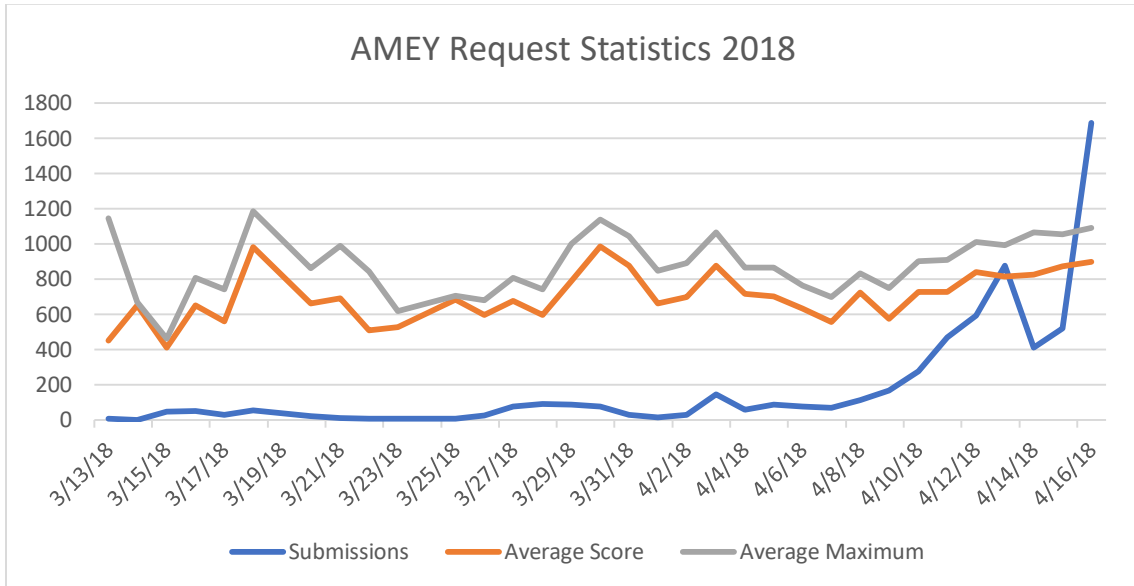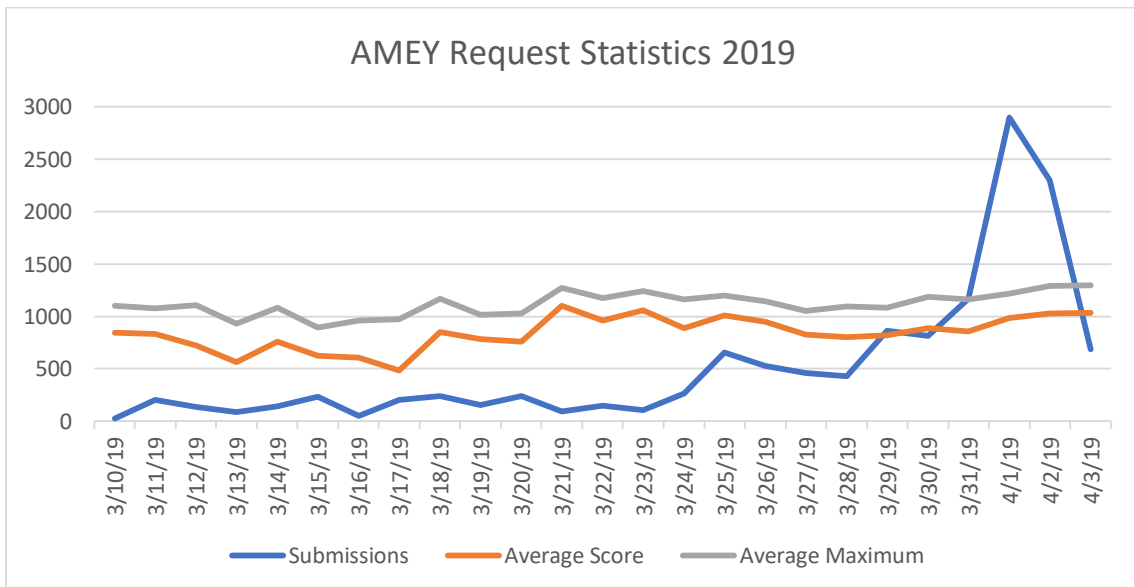

Figure 2. AMEY statistics for 2019 cohort showing number of submissions along with both the average score and average maximum achievable scores (of elements submitted).

4. Discussion and evaluation

The use of automated assessment and support in teaching such a large cohort was seen as essential given the relatively low staffing resources (one academic and one teaching associate) assigned to the module. Using this in part to meet the university requirements for formative feedback was seen to be promising and the results both in feedback from learners and anecdotal evidence from staff on the positive impact on their workload bears this out. Automated feedback and/or assessment is by no means a "one size fits all" option, the domain must be suitable ideally containing absolute right or wrong answers which can be evaluated.

Even within the context of this assessment on databases where there is a correct answer for each and every item use of automated assessment is far from perfect. When marking manually consideration may be given for minor structural or presentational issues, for example a spelling mistake or the inclusion of spaces in an element where there should be none. In such cases it's common practice to assign some penalty to the error but still mark the overall answer for general correctness against a rubric. Without developing much "cleverer" automated systems which could detect these sorts of minor mistakes automated marking is completely "black and white". For example, a mis-named table in the database would result in that table scoring 0 and all the expected components in that table also being scored 0 even if the individual fields and structure were themselves correct.

Using the automated approach to provide early access as formative feedback but on the exact same basis on which summative marking is performed addresses this problem. Learners on submission would regularly discover that an entire table was being scored zero and be able to investigate, determining that naming was at fault and not just fixing that for the table in question but inferring that they should pay close attention to naming conventions throughout. Common errors in this way are identified and overcome by the learner problem solving themselves as to why points were missing from the scores.

The provision of AMEY and the feedback it provided were appreciated by learners as shown in their end-of-year module evaluations (at the time of this paper only evaluations for 2018 are available) which were overwhelmingly positive and included comments that AMEY "saved my project" and they "wish[ed] everyone did this". The negative comments in 2018 revolved around accessibility from off-campus (fixed for 2019) and too much information in the output. Analysis of submissions along with anecdotal evidence based on student email queries shows a distinct group of "early adopters" who made continued use of AMEY to refine their project as well as a number of "late adopters" who it can be assumed started their projects much nearer to submission. In 2018 an optional anonymous survey of students was performed which received 90 responses. Of those respondents who had used AMEY 100% either agreed or strongly agreed that "[AMEY was] useful in preparing my coursework" and 98.3% that they were "glad such as system was available". Beyond just getting the coursework to submissions standard 86.2% agreed or strongly agreed that their "understanding of SQL increased through finding and fixing errors [with AMEY]".

5. Conclusions and future work

In this paper we have presented the findings of provisioning an automated feedback tool which provided learners early access to a subset of their summative assessment criteria for project work in a large first-year computer science class on databases. Originally intended purely to support marking, opening the tool up in part to learners themselves led to widespread use of the tool, positive evaluations, reduced workload for teaching staff, and hopefully a greater understanding of the content through pro-active problem solving. Although far from perfect we believe the solution was highly beneficial in terms of this module and also shows the potential for a new kind of formative-summative pairing with automated assessment in future.

The next steps will be to continue to refine AMEY in terms of database modules and consider other modules or areas in which such an approach could be applied. The refinement to AMEY will focus in large part of the user experience and trying to minimize the amount of data presented to the user where there is no issue for them to address, focusing on problem areas only. Research will also take place into how any "fuzzy" approaches could be incorporated to identify things such as minor typos and correcting them for assessment with just a minor deduction to the score.

References

Blumenstein, M., Green, S., Nguyen, A. & Muthukkumarasamy, V. (2004), "Game: A generic automated marking environment for programming assessment", *Information Technology: Coding and Computing, 2004. Proceedings. ITCC 2004. International Conference on*IEEE, pp. 212.

Gerdes, A., Heeren, B., Jeuring, J. & van Binsbergen, L.T. (2017), "Ask-Elle: an adaptable programming tutor for Haskell giving automated feedback", *International Journal of Artificial Intelligence in Education,* vol. 27, no. 1, pp. 65-100.

Heckman, S. & King, J. (2018), "Developing Software Engineering Skills using Real Tools for Automated Grading", *Proceedings of the 49th ACM Technical Symposium on Computer Science Education*ACM, pp. 794.

Joy, M. & Luck, M. (1998), "Effective electronic marking for on-line assessment", *ACM SIGCSE Bulletin*ACM, pp. 134.

Leckey, J. & Neill, N. (2001), "Quantifying quality: the importance of student feedback", *Quality in Higher Education,* vol. 7, no. 1, pp. 19-32.

Milanova, A., Cutler, B., Holzbauer, B., Maicus, E., Breese, S., Dinella, E. & Peveler, M. (2018), "Program Analysis Tools in Automated Grading of Homework Assignments", *Proceedings of the 49th ACM Technical Symposium on Computer Science Education*ACM, pp. 1090.

Quinton, S. & Smallbone, T. (2010), "Feeding forward: using feedback to promote student reflection and learning--a teaching model", *Innovations in Education and Teaching International,* vol. 47, no. 1, pp. 125-135.

Rowe, A.D. & Wood, L.N. 2008, "Student perceptions and preferences for feedback", *Asian Social Science*, vol. 4, no. 3, pp. 78-88