ŌTAUTAHI CHRISTCHURCH AOTEAROA NEW ZEALAND
35TH AUSTRALASIAN ASSOCIATION FOR ENGINEERING EDUCATION ANNUAL CONFERENCE
8-11 DEC 2024
THE ENGINEER AND THE WORLD

Student perspectives on delayed summative assessment

Raveen de Silva; and Jake Renzella.
UNSW Sydney
Corresponding Author Email: r.desilva@unsw.edu.au

ABSTRACT

CONTEXT

Algorithm design and analysis is fundamental in computer science and software engineering, serving as a foundation for further theoretical study while also developing analytical and problem-solving skills. Employers value these skills particularly highly, as exemplified by the prevalence of algorithmic problems in technical interviews for software engineering roles. However, this subject area can pose multifaceted challenges to students, as it is highly conceptual, requires mathematical maturity and involves written communication rather than programming. At UNSW Sydney, algorithm design and analysis is jointly taught to second- or third-year computer science and software engineering students and to postgraduate coursework students in IT.

PURPOSE OR GOAL

This paper evaluates students' experiences of a redesign of the standard algorithms course at UNSW Sydney. Many students struggle to grasp and apply the threshold concepts, and as a result the course is perceived as being difficult and esoteric. We identified shortcomings in the interactivity, guidance and feedback involved, which are especially important for novice learners, and changed the assessment structure in line with the backward design philosophy to foster greater student achievement and engagement rather than purely strategic learning.

APPROACH OR METHODOLOGY/METHODS

Students in two offerings of the course were surveyed regarding their experiences of learning and assessment, once under a traditional series of assignments, and once under a portfolio assessment. Students were asked a range of questions on self-evaluation of their fulfilment of the course aims, assessment experiences, guidance and feedback received, and grade expectations. Comparison of these survey responses allows us to evaluate the effectiveness of the course redesign and its impact on student experience.

ACTUAL OR ANTICIPATED OUTCOMES

Students in the redesigned offering reported greater achievement of the course aims, greater satisfaction with assessment and staff support, and reduced grade anxiety.

CONCLUSIONS/RECOMMENDATIONS/SUMMARY

Delayed summative assessment has been implemented successfully in some computing courses across multiple institutions in Australia, promoting deeper learning and more constructive student experience. This study shows its effectiveness in improving student experience of a very large and technical course in algorithm design and analysis, which has historically been rather ineffective at its stated aims and difficult for students to approach.

KEYWORDS
Computing education, learner-centred teaching, algorithms, delayed summative assessment

Background

Algorithm design and analysis

Algorithm design and analysis is a fundamental area of computer science and software engineering coursework. While introductory courses train students to write code that implements simple processes and data structure operations, a later course is typically reserved for students to learn algorithm design paradigms such as divide-and-conquer, greedy and dynamic programming, as well as the associated algorithm analysis. Such a course may also serve as students' first introduction to intractable problems. The best-known text in this area is *Introduction to Algorithms* (Cormen et al., 2022). At UNSW Sydney, algorithm design and analysis is jointly taught to second- or third-year computer science and software engineering students and to postgraduate coursework students in information technology, who have completed prior courses on data structures and discrete mathematics.

Students develop their problem-solving skills through algorithm design, and also think critically and practice clear and precise written communication in analysing the correctness and efficiency of algorithms. Most famously, algorithms problems are ubiquitous in technical interviews for software engineering roles (Aziz et al, 2015). Although this practice is sometimes criticised, such as for inauthenticity and inequity (Behroozi et al., 2019), it can be expected that these skills will only grow in importance with the greater availability of large language models which perform well in routine programming tasks but struggle with complex reasoning (Prather et al., 2023).

Challenges

Students often report that algorithm design and analysis is a difficult course. It is highly abstract, particularly as correctness is established *a priori* rather than from a suite of concrete test cases. It also requires mathematical maturity, wherein the common procedural understanding of mathematical concepts is insufficient, most notably proof by induction as highlighted by Baker (1996). Finally, written communication is often an under-developed skill in computer science education, which also poses particular challenges to students with limited English proficiency.

Traditional assessment

Aside from exams, the most common forms of assessment in algorithms courses are problem-solving assignments. These may require students to prepare a written response to each problem, as was historically the practice at UNSW Sydney, or write a program which efficiently implements their algorithm, usually in a prescribed programming language. Over several years of conducting such written assignments, we identified a number of concerns, particularly in assessment incentives, opportunities for feedback and guidance, and grade anxiety.

Assessment incentives

Ramsden (2003) and Blumberg (2008) discuss various approaches to learning, identifying that students often engage in a strategic manner, seeking primarily to maximise assessment results and sometimes failing to make deeper connections. In the traditional structure, students were observed to have limited acquisition of course learning outcomes, and sometimes a very poor grasp of subtopics that were not covered in assignment questions.

Feedback and guidance

This course is very large (in excess of 500 students), so marking assignments was a highly structured process conducted by a large team of sessional staff. To maintain the integrity of the assessment, staff did not help students to attempt the assignment questions, but rather aimed provided guidance on thematically similar problems. These classes and resources were however greatly underutilised due to the lack of direct connection to assessment. Feedback was limited to individual and course-wide comments explaining the marking of assignment submissions, and by the time these could be returned to students, classes had moved on to other topics so students did not have an opportunity to implement these suggestions.

Grade anxiety

Assignment questions were marked numerical according to marking criteria, with auditing to ensure consistency. Students often reported anxieties about marking quality and consistency, perhaps due to a lack of transparency before the submission deadline and also the lack of a relationship between student and assessor. Students had difficulty in assessing the quality of their submission, leading to surprise and disappointment upon the release of marks. Students were also incentivised to submit work that they knew had no merit in the hope of acquiring some marks, which was both an unethical practice and a drain on marking resources.

Portfolio assessment

Portfolio assessment is an alternative to traditional assignments, in which students produce a body of work across one or more teaching periods, to exhibit both the product and the process of learning. While it has been commonplace in humanities and artistic disciplines for some time, portfolio assessment is increasingly popular in STEM, with benefits including the promotion of self-reflection as discussed by Fielke and Quinn (2009, 2011).

Delayed summative assessment

Portfolio assessment is most naturally applied to subject areas with few discrete tasks each worked on over a long time period, which is not typical of general-audience computer science courses. However, Renzella and Cain (2017) pioneered a novel system of task-oriented portfolio assessment, where students attempt a suite of competency-based assessments, and receive frequent formative feedback to guide them through their learning. This was implemented using the Doubtfire Learning Management System, which has been described at depth in literature (Cain et al., 2020). In summary, tasks are classified into grade levels with reference to the structure of observed learning outcomes (SOLO) taxonomy developed by Biggs and Collis (1982), where Pass and Credit tasks require uni- and multi-structural understanding, while Distinction and High Distinction tasks focus on relational and extended abstract thinking. Students attempt tasks up to their chosen target grade, which they can change at any time. Each task is then a discrete activity, which progresses through states from “Not Started” to “Complete” as the student makes a first attempt and iterates upon it with feedback until a satisfactory standard has been achieved. Instructors provide students with several rounds of formative feedback in the form of text or audio comments, which students must address in resubmissions in order to complete the task. Subjective assessment can be combined with autotesting if desired.

At the end of the teaching period, students submit a learning summary report to reflect on their attainment of the course learning outcomes, which together with the aggregation of their completed tasks is the main basis for tutors to award a portfolio mark. In this way, a summative portfolio assessment can be comprised of individual formative components. Note that due to timing constraints, our implementation of this system did not require the learning summary report, and marks were calculated directly from task completion without tutors’ input.

Delayed summative assessment using the Doubtfire platform has been implemented in several introductory programming courses in Australia, for example as documented by Cain and Woodward (2013), Cain and Babar (2016) and Cain et al. (2018), but little research has been previously been published on its applicability to other courses in computing or other disciplines.

Redesign

The algorithm design and analysis course at UNSW Sydney was redesigned between 2023 and 2024 to replace the traditional series of assignments with a portfolio, using the Doubtfire LMS to implement delayed summative assessment. Tasks were developed with the aim of covering the full breadth of content, so that students would develop comprehensive knowledge and skills.

Pass and Credit tasks focused on recollection, understanding and explanation of algorithmic principles and direct application of known algorithms to problems, whereas Distinction and High Distinction tasks involved more independent problem solving as well as precise analysis and proofs. This also coincided with renewed clarity in the aims of the course, as expressed in the learning outcomes, with more explicit focus on algorithm analysis and written communication.

The remainder of this paper discusses surveys issued to cohorts before and after the implementation of this course redesign, in order to evaluate its effectiveness in resolving concerns present from the traditional structure and identify areas for improvement.

Method

This course was undertaken by 685 students in the third term of 2023 (23T3) with traditional assignments, and by 577 students in the first term of 2024 (24T1) with delayed summative assessment as described in the previous section. At the end of each of these terms, students were invited to participate in an anonymous online survey. Both surveys consisted of the same twenty Likert items, in four categories as below, each rated on a five-point scale. Students were informed that the survey was voluntary, their participation or non-participation could not be identified, and it was in no way connected to assessment in the course. There were 67 responses to the survey issued to the 23T3 cohort (52 complete, 15 incomplete), and 60 responses to that issued to the 24T1 cohort (49 complete, 11 incomplete). It is not possible to determine whether or not the sample of students who took part in either survey is representative of the enrolled cohort.

Course learning outcomes

The first four items address 24T1 outcomes, which were clarified from similar outcomes in 23T3.

Code	Item
LO1	I have a thorough understanding of algorithm design techniques.
LO2	I am able to design algorithms to solve computational problems.
LO3	I am able to analyse the correctness and efficiency of algorithms.
LO4	I am able to communicate clearly about algorithmic ideas.
LO5	My skills meet the expectations of prospective employers.

Table 1: Items on attainment of learning outcomes.

Assessment design

The next six items relate to the assessment design.

Code	Item
AD1	Assessment tasks accurately measured my algorithm design and analysis skills.
AD2	Assessment tasks accurately measured my communication skills.
AD3	Completing assessment tasks was an important part of my learning.
AD4	Assessment tasks directed me to learn the breadth of course material.
AD5	Assessment tasks directed me to learn course material in depth.
AD6	Assessment methods reflected the effort I put into the tasks.

Table 2: Items on assessment design.

Feedback

The next five items relate to feedback from staff.

Table 3: Items on feedback.

Code	Item
FB1	Staff provided useful guidance that was specific to me.
FB2	Staff provided accurate feedback on my work.
FB3	Staff provided timely feedback on my work.
FB4	Staff provided constructive feedback on my work.
FB5	Staff provided feedback that helped me in my work for the rest of the term.

Grade expectations

In both terms, some respondents completed the survey before and some after the final exam, but both surveys were closed before exam marks and overall course marks were published.

Table 4: Items on grade expectations.

Code	Item
GE1	I have a clear expectation of approximately what mark I will receive.
GE2	My expected mark is satisfactory to me.
GE3	My expected mark reflects the quality of my work.
GE4	My expected mark reflects the effort put into my work.

Results

In each table below, the responses to each item are split into two columns, with the 23T3 cohort on the left (shaded in orange) and 24T1 on the right (shaded in blue). The quartiles are reported using the standard enumeration (1 = strongly disagree, …, 5 = strongly agree).

Course learning outcomes

Table 5: Responses to items on learning outcomes.

Response	LO1		LO2		LO3		LO4		LO5	
Strongly disagree	7	1	4	1	5	0	5	0	3	2
Somewhat disagree	8	4	8	3	8	0	9	5	10	7
Neither agree nor disagree	12	3	5	6	8	9	13	11	27	15
Somewhat agree	26	42	40	35	35	35	27	30	18	26
Strongly agree	14	10	10	15	11	16	13	14	9	10
Q1	3	4	3	4	3	4	3	3	3	3
Q2	4	4	4	4	4	4	4	4	3	4
Q3	4	4	4	4.5	4	5	4	4	4	4

Assessment design

Table 6: Responses to items on assessment design.

Response	AD1		AD2		AD3		AD4		AD5		AD6	
Strongly disagree	4	2	4	2	2	3	4	1	5	4	6	3
Somewhat disagree	9	6	6	3	4	3	3	1	3	2	8	3
Neither agree nor disagree	3	3	7	2	4	1	2	1	3	3	8	6
Somewhat agree	29	17	27	26	19	17	29	25	27	20	23	18
Strongly agree	12	23	13	18	28	27	19	23	19	22	12	21
Q1	3	4	3	4	4	4	4	4	4	4	3	4
Q2	4	4	4	4	4	5	4	4	4	4	4	4
Q3	4	5	4	5	5	5	5	5	5	5	4	5

Feedback

Table 7: Responses to items on feedback.

Response	FB1		FB2		FB3		FB4		FB5	
Strongly disagree	4	3	4	1	4	2	3	2	5	2
Somewhat disagree	7	1	7	5	8	15	7	2	5	5
Neither agree nor disagree	10	4	8	4	8	8	6	6	14	8
Somewhat agree	20	18	26	16	20	14	25	21	20	18
Strongly agree	11	24	7	24	12	11	11	19	8	17
Q1	3	4	3	4	3	2	3	4	3	3
Q2	4	4	4	4	4	3.5	4	4	4	4
Q3	4	5	4	5	4	4	4	5	4	5

Grade expectations

Table 8: Responses to items on grade expectations.

Response	GE1		GE2		GE3		GE4	
Strongly disagree	7	4	12	4	5	2	10	3
Somewhat disagree	11	3	16	6	9	4	9	5
Neither agree nor disagree	15	5	7	12	13	8	11	8
Somewhat agree	17	25	13	18	19	20	16	16
Strongly agree	2	12	4	9	6	15	6	17
Q1	2	4	2	3	2	3	2	3
Q2	3	4	2	4	3	4	3	4
Q3	4	4	4	4	4	5	4	5

Discussion

For each item, results from the two cohorts were compared using a Wilcoxon rank sum test, with one-sided significance threshold α = 0.05.

Learning outcomes

Results for all five of these items were significantly higher in the 24T1 cohort, most notably LO1 (U = 1629.5, $p = 0.022$) and LO3 (U = 1535.5, $p = 0.006$).

As LO1 ("I have a thorough understanding of algorithm design techniques") relates to more fundamental knowledge, this difference may be attributed to its inclusion of lower grade assessment tasks that students are incentivised to complete and engage with feedback on, where previously this was not directly assessed in assignments.

LO3 also saw substantially higher results. In the past, assignment questions would typically require students to solve an algorithm design problem in order to exhibit their algorithm analysis skills, but the inclusion of tasks where students analysed a given algorithm proved impactful, perhaps particularly for those with less developed problem solving skills.

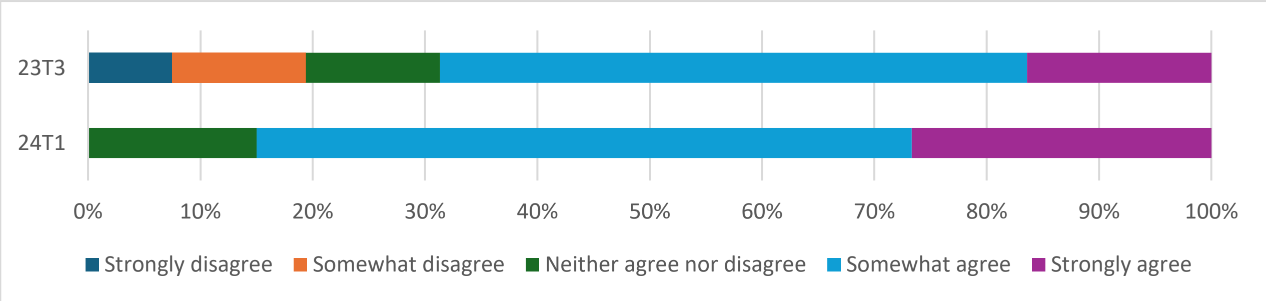

Figure 1: Responses to LO3 "I am able to analyse the correctness and efficiency of algorithms.".

Assessment design

Results were significantly higher in the 24T1 cohort for AD1 (U = 1117, $p = 0.014$), AD2 (U = 1153, $p = 0.023$), AD4 (U = 1211.5, $p = 0.050$) and AD6 (U = 1086, $p = 0.009$).

24T1 students reported that their skills in algorithm design and analysis and in communication were more accurately assessed, they were assessed on more of the full breadth of course material, and their effort was more accurately reflected in assessment. These results are in line with our hypotheses, as the ability to respond to feedback and resubmit made students focus on communicating sincerely rather than maximising marks, the larger number of portfolio tasks allowed greater coverage of the syllabus, and assessment results were clearly tied to student commitment, with relatively little fruitless effort.

Feedback

One challenge we faced in implementation was in marking submissions promptly, particularly as too many tasks were assigned at the start of the term, and the sessional team had to become familiar with the whole suite of tasks. This resulted in a modest but not statistically significant decline in satisfaction with timeliness of feedback, shown in item FB3 (U = 1178, $p = 0.200$). We expect this will improve in repeat offerings.

However, students reported significantly greater satisfaction with the quality of feedback in 24T1. The term-long relationship between student and lab demonstrator enabled more personalised, accurate and constructive feedback, seen in FB1 (U = 848.5, $p < 0.001$), FB2 (U = 824.5, $p < 0.001$) and FB4 (U = 1024, $p = 0.025$) respectively. The shorter feedback cycle was particularly important in ensuring continuity of feedback, as reflected in FB5 (U = 993.5, $p = 0.016$).

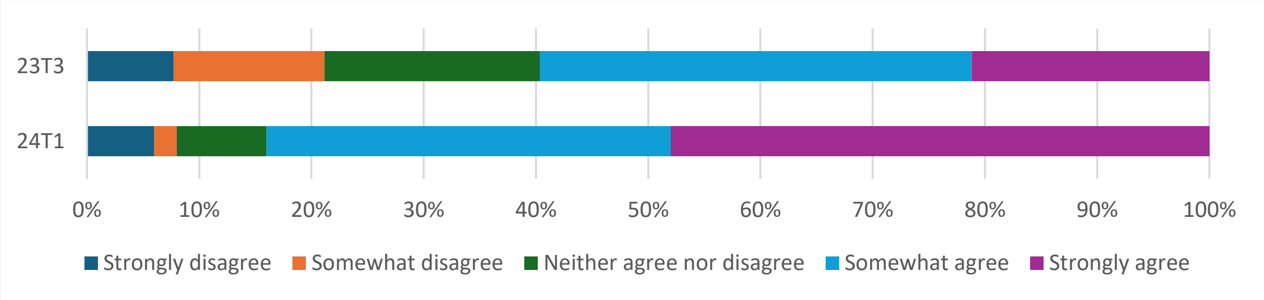

Figure 2: Responses to FB2 “Staff provided useful guidance that was specific to me.”.

Grade expectations

All four items had significantly higher results in 24T1, aligning with anecdotal observations that most students in a Doubtfire assessment achieve their target grade, and existing research showing that students are better able to reflect on their work in portfolio assessments (Fielke & Quinn, 2009, 2011). It is particularly notable that these effects were achieved even though the students did not engage in a reflective activity when assembling the final portfolio.

The most improvement was seen in GE1 ($U = 719.5$, $p < 0.0001$), where we observed that students had much less uncertainty about grades, which is a common source of anxiety and dissatisfaction. This may also reflect greater understanding of what constitutes good work in an exam setting, due to the constant feedback provided during the teaching term.

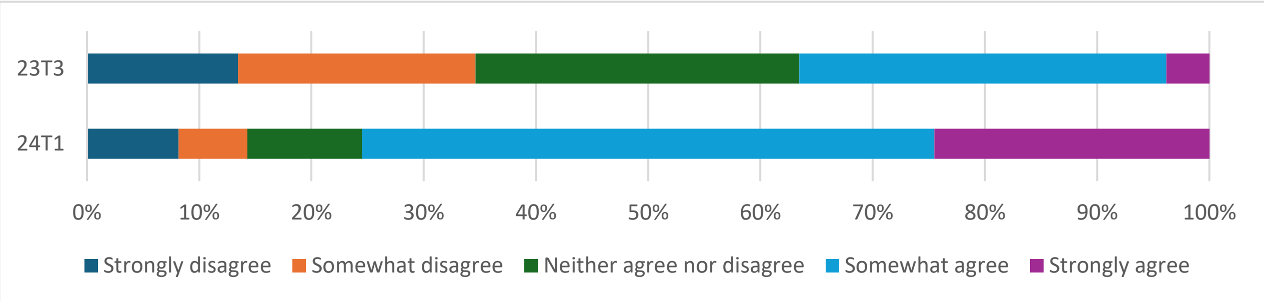

Figure 3: Responses to GE1 “I have a clear expectation of approximately what mark I will receive.”.

GE2 ($U = 829$, $p < 0.001$) reflects that students were more satisfied with their expected mark in 24T1. While some students in both cohorts indicated that they hoped for higher marks at the start of term, GE3 ($U = 876$, $p = 0.002$) and GE4 ($U = 823.5$, $p < 0.001$) show that the 24T1 cohort had a clearer understanding of why their expectations were or were not fulfilled.

Conclusion

It is clear that delayed summative assessment had a significant positive impact on student experience in this course. The new assessment structure supported alignment between learning outcomes, assessment and learning activities, which led students to be more confident that they had met the aims of the course, improved perceptions of assessment accuracy and fairness, facilitated constructive engagement with feedback and reduced grade anxiety. Further refinements in this course will revise existing tasks and lower workload for students and staff.

We look forward to the wider adoption of task-oriented portfolio assessment outside of introductory programming courses.

References

Aziz, A., Lee, T. -H., & Prakash, A. (2015). *Elements of Programming Interviews: The Insider’s Guide*.

Baker, J. D. (1996). *Students' Difficulties with Proof by Mathematical Induction*. Paper presented at the Annual Meeting of the American Educational Research Association, New York, NY.

Behroozi, M., Parnin, C. & Barik, T. (2019). *Hiring is Broken: What do Developers Say About Technical Interviews?* Paper presented at the 2019 IEEE Symposium on Visual Languages and Human-Centric Computing, Memphis, TN.

Biggs, J. B., & Collis, K. F. (1982). *Evaluating the Quality of Learning: The SOLO Taxonomy*. New York, NY: Academic Press.

Blumberg, P. (2008). *Developing Learner-Centred teaching: A Practical Guide for Faculty*. San Francisco: Jossey-Bass.

Cain, A., & Babar, M. A. (2016). *Reflections on applying constructive alignment with formative feedback for teaching introductory programming and software architecture*. Paper presented at the 38th International Conference on Software Engineering, Austin, TX.

Cain, A., Grundy, J., & Woodward, C. J. (2018). Focusing on learning through constructive alignment with task-oriented portfolio assessment. *European Journal of Engineering Education, 43*(4), 569-584.

Cain, A., Tubino, L., & Krishnan, S. (2020). Using Technology to Enable a Shift from Marks to Outcomes-Based Assessment. In Bearman, M., Dawson, P., Ajjawi, R., Tai, J., & Boud, D. (Eds.), *Re-imagining University Assessment in a Digital World* (pp. 249-245). Springer.

Cain, A., & Woodward, C. J. (2013). *Examining Student Reflections from a Constructively Aligned Introductory Programming Unit*. Paper presented at the Fifteenth Australasian Computing Education Conference, Adelaide, SA.

Cormen, T. H., Leiserson, C.E., Rivest, R.L., & Stein, C. (2022). *Introduction to Algorithms* (4th ed.). Cambridge, MA: MIT Press.

Fielke, J., & Quinn, D. (2011). *Improving student engagement with self-assessment through ePortfolios*. Paper presented at the Australasian Association for Engineering Education Annual Conference, Fremantle, WA.

Fielke, J. M., & Quinn, D. (2009). *Supporting self-assessment using e-portfolios*. Paper presented at the Australasian Association for Engineering Education Annual Conference, Adelaide, SA.

Prather, J., Denny, P., Leinonen, J., Becker, B. A., Albluwi, I., Craig, M., Keuning, H., Kiesler, N., Kohn, T., Luxton-Reilly, A., MacNeil, S., Petersen, A., Pettit, R., Reeves, B. N., & Savelka, J. (2023). *The Robots are Here: Navigating the Generative AI Revolution in Computing Education*. Paper presented at the 28th Annual Conference on Innovation and Technology in Computer Science Education, Turku.

Ramsden, P. (2003). *Learning to teach in higher education* (2nd ed.). London: Routledge.

Renzella, J., & Cain, A. (2017). *Supporting better formative feedback in task-oriented portfolio assessment*. Paper presented at the 2017 IEEE International Conference on Teaching, Assessment and Learning for Engineering, Hong Kong.

Ethics Approval

Ethics approval for the use of student survey data in this study was granted by the UNSW HREAP Executive (reference number iRECS5021).

Acknowledgements

The authors would like to thank the students who provided their feedback in the surveys, and especially the many sessional staff who contributed to the success of the course redesign.