ASCILITE 2025

Future-Focused:

Educating in an Era of Continuous Change

Scaffolding students' learning of introductory programming in online higher education by personalised formative assessments

S. M. Masud Karim
UniSA STEM, University of South Australia

Siamak Mirzaei, Shekh Nisar Hossain, Rhodora Abadia
UniSA Online, University of South Australia

Aminul Islam
Computer Science and Engineering Discipline, Khulna University, Bangladesh

This paper presents the results of a study to investigate the application of personalised assessment in supporting students' learning of introductory programming. The participants of the study were university students majoring in information technology and data analytics. The students in the study were added to a Moodle course called Computer Programming in Python, where they were allowed to take a test implemented using personalised assessment and another test implemented using traditional assessment (i.e., timed, fixed length and randomised ordering) as many times as they intend. The tests consist of mixed types of questions ranging from single word answers, True-False to multiple choice questions with three, four or five possible answers with one correct answer. The questions for the test using personalised assessment are divided into three categories according to the level of difficulty (easy, medium and hard). Based on these three categories, students' abilities were estimated as Beginner, Intermediate, Advanced and Expert and detailed personalised feedback were provided instantly highlighting the concepts and issues in their attempts and linking with the course contents and course objectives. The results of the study show that the personalised assessment significantly outperforms traditional assessment with respect to usability, effectiveness and immediate quality feedback.

Keywords: Personalised assessment, formative assessment, introductory programming, Python programming, students' learning, feedback, mixed method.

1. Introduction

The term 'introductory programming', also known as CS1, refers to an information technology (IT) course for novice students that typically covers problem-solving techniques, basic programming concepts, the syntax and semantics of a programming language, and the use of the programming language to formulate solutions (Medeiros et al., 2019). Weekly formative assessments scaffold students' learning with guidance and feedback, helping students to complete the summative assessments, and as such to achieve course objectives (COs). Although advanced methods and tools are used for teaching and learning (Resnick et al., 2009; Mohorovicic & Strcic, 2011), dropout and failure rates remain high in introductory programming courses (Watson & Li, 2014; Bennedsen & Caspersen, 2007; Bennedsen & Caspersen, 2019).

Quality feedback can enhance students' learning and reduce failure rates. Hattie (2009) ranked feedback among the top 10 influences on students' achievement of the 138 influences in over 800 meta-reviews. Quality feedback is regarded as a major contributor to improving students' learning (Hattie & Timperly, 2007). However, the feedback should be timely and meaningful (Barker, 2010). Providing timely feedback is considered as one of the key principles of designing for learning (Stone & Crawford, 2020; UCF, 2011).

Personalised learning correlates educational concepts with students' interests, abilities, and experiences. It adopts tailored curriculum, activities, and assessments to individual needs, rather than enforcing a 'one size

fits all' model on students (UNSW eLearning, 2014). Personalised assessment, a core feature of personalised learning, offers greater flexibility for a personalised experience for each student and helps students to recognise their strengths and scopes for improvement. Assessment also serves as a natural channel for delivering feedback within personalised learning environments (Evans, 2013).

Computerised adaptive tests (CATs) are commonly used to measure knowledge acquisition but mostly offer only a final score or a relative score with respect to all participants with no or little feedback. Although formative assessment has been extensively studied in elementary and secondary education, implementing it for personalised learning in higher education is still limited (Gierl et al., 2018). The true potential of personalised assessment can be achieved by utilising it as formative assessment, where quality feedback will scaffold students' learning and build their pathway to grow. Students' learning can be facilitated by personalised formative assessment, especially when task-specific, immediate feedback is provided to students (Black & William, 2009).

This study develops a prototype of a personalised formative assessment using CAT to scaffold students' learning in an introductory Python programming course in online higher education. The participants of the study were enrolled as students in a mock course, where they sat for tests using both traditional assessment and personalised assessment in unlimited attempts and completed the 3 surveys. Findings indicate that personalised assessment outperforms traditional assessment in usability, effectiveness and quality feedback.

2. Personalised Assessments

In IT courses, formative assessments can be used throughout the courses to scaffold students towards the major summative assessments. Personalised assessment can be utilised as continuous formative assessments as well as the major summative assessment. This provides a tailored assessment experience for each student and will help all students develop their skills through understanding what they can do, the things they need to work on, and their next steps.

Students may encounter various difficulties due to the students' weaknesses (including learning difficulties), perceptions, and interests. Administering a common set of assessments for all students of a course having the same requirements, activities, and timeframe may introduce issues and thus impact students' learning. Every student is different; hence we may need to apply empathy and make flexible arrangements in assessments to allow students to achieve COs. The flexibility may range from adopting personalised selection for online assessment modes, where student can select oral final viva or written final exam as their final major summative assessments to customising personalised assessment environments, e.g., viewing the assessment questions in a particular background colour and layout, changing texts in preferred font and size and reading the questions in a preferred alternative language for better understanding of the requirements etc.

The most popular and state of the art approach for personalised assessment is realised using adaptive technologies. Tests having a fixed question sequence or randomised question selection may present students with questions which they consider problematic. Students may become discouraged when encountering overly difficult questions, but questions that are too easy can also cause disengagement due to lack of challenge (Cisar et al., 2016). In a CAT, questions are selected based on the student's response to the previous questions. Students can select the difficulty level of the questions they want to start exploring and move forward through more challenging questions incrementally. This approach will boost students' motivation and confidence and enhance their learning.

To the best of our knowledge, the application of personalised assessment in online higher education is limited. The adaptive approaches, where randomised or formula-based questions are developed, are done to mainly to mitigate academic misconducts, they are not based on students' needs or interests. The closest approach that has been implemented with respect to personalised assessment environment setting, when extra times are allocated and/or minor adjustments in the assessment requirements are allowed for different students based on their special needs mentioned in their access plans. Another practice can be considered as closely associated with personalised assessments, where personalised interventions have been implemented. For instance, sending emails to students using OnTask+ (Pardo et al., 2018) based on their assessment status and

outcomes, e.g., congratulating students for showing outstanding performances in an assessment, reminding students to submit their assessments and so on.

3. Background of CATs

CAT is among the most widely adopted technologies for enabling personalised assessment. However, CAT has traditionally been designed for determining the proficiency for awarding certifications e.g., pathologists and laboratory professionals (ASCP BOC, n.d.), armed services (ASVAB, n.d.), evaluating knowledges and/or qualifications like GMAT (GMAC, 2025), Naplan (ACARA, n.d.), or testing language proficiency such as English (Jacobsen et al., 2011; Lilley et al., 2004; Olea et al., 2011; Olea et al., 2012), Japanese (TTBJ, n.d.), second national language – either Dutch or French (Schonenberg et al., 1993).

Chatzopoulou & Economides (2010) presented Programming Adaptive Testing (PAT), a web-based adaptive testing system for assessing students' programming knowledge on 'Glossa' (a pseudo-language which can be best described as a Greek translation of PASCAL language) in an introductory programming course in a high school in Greece. The system adapted question difficulty based on student responses—presenting harder questions after correct answers and easier ones after incorrect responses. Simple questions tested basic knowledge, while challenging questions assessed students' ability to apply what they had learned to novel situations. Students completed a customised 30-question assessment, and PAT categorised them into one of three competency levels: High, Moderate and Low. However, PAT only tested beginner level of programming. In addition, it is not for formative assessment, and it did not provide formative feedback to guide the learning.

Cisar et al. (2016) applied a CAT in knowledge evaluation of the programming language C++ of the first-year college students at a polytechnic college in Republic of Serbia majoring in computer engineering. The study compared hand scoring done for the paper-and-pencil test (PPT) version and automatic scoring by computer for the CAT and showed that the students who had worked on a CAT achieved a higher average score than the students who did the PPT version.

The eDia online assessment system has been developed to provide regular diagnostic information in 3 main domains of education: reading, mathematics, and science, from the beginning of schooling to the end of the 6 years of primary education (Csapo & Molnar, 2019). It uses formative assessment with 2 types of feedback i.e., (i) immediate feedback and (ii) situational feedback based on comparative standards, accessible only following the main assessments as diagnostic assessment.

Gierl et al. (2018) lists 5 requirements for a computerised test to be used for personalised learning:

- **Self-assessment**: It should enable students to conduct their own self-evaluations,
- **On-demand**: Students must be able to access it anytime, as often as necessary,
- **Immediate feedback**: Students must get instant feedback showing if their answer to each question is right or wrong with necessary explanation,
- **Big test items**: Students must be able to access a range of content-aligned test items created to gauge targeted learning outcomes.
- **Parallel test forms**: Students should be able to retake equivalent tests, with different questions but measuring the same skills, after modifying their study approach.

The theories, algorithms and techniques of different types of assessments in personalised learnings are summarised in a recent systematic review by Ihichr et al. (2024).

4. Experimental Design

To conduct the experimental study, a mock course titled *Computer Programming in Python* is created on Moodle LMS installed at a private website. The course contents are presented into 8 sections with each section covers a single topic – (i) Instruction to Python, (ii) Console Input and Output, (iii) Expressions and Statements, (iv) Control Structures – Selection, (v) Control Structures – Repetition, (vi) List and String, (vii) Functions, and (viii) File Input and Output. The course is designed with the following COs:

CO1. Implement programs to solve problems.

CO2. Choose appropriate conditional and iteration constructs for a given programming task.
CO3. Analyse and explain the behaviour of programs.
CO4. Test and debug algorithms and programs and describe strategies that are useful in debugging.
CO5. Apply good programming practices, specifically, coding style standards and source code documentation.

The course contents and the COs are mainly created to demonstrate the personalised feedback of the personalised assessment. The traditional assessment is implemented using the Moodle Quiz tool, whereas the personalised assessment is implemented by using a modified version of Moodle plugin Adaptive Quiz (Franco & Potenko, 2022) following the requirements of Gierl et al. (2018) and it is called *persona*.

Two separate question banks are created for the online assessments: one for traditional assessment and a separate one for persona. There are 3 levels of difficulty of the questions in persona: easy, medium and hard. Using revised Broom's taxonomy (Krathwohl, 2002), the factual and conceptual knowledge is placed with the cognitive process of remember and understand to get easy question items, procedural knowledge is placed with understand and apply to obtain medium question items and finally all dimension of knowledge including metacognitive knowledge is tied with apply and analyse to compose the hard question items. For calibrating the items for this study, expert calibration (Lilley, 2007) is used.

For question selection and test administration, we use testlet-based CAT instead of fully adaptive testing, where item-based CAT is used (Eggen, 2007). We have one testlet for each difficulty level with a fixed number of questions to be posed during a test. First the students will get random questions from easy testlet, then random questions from medium testlet if they do well i.e., pass in easy testlet, and finally random questions from hard testlet if they do well, i.e., pass in medium testlet. We set 50% score as the default pass threshold for each level, but this can be modified as required. The test using persona ends when students fail to pass a level or complete the final level. The ability of the students will be estimated as 0, 1, 2, and 3, where level 0 is for Beginner, 1 denotes Intermediate, 2 means Advanced and 3 refers to Expert, as shown below in Figure 1.

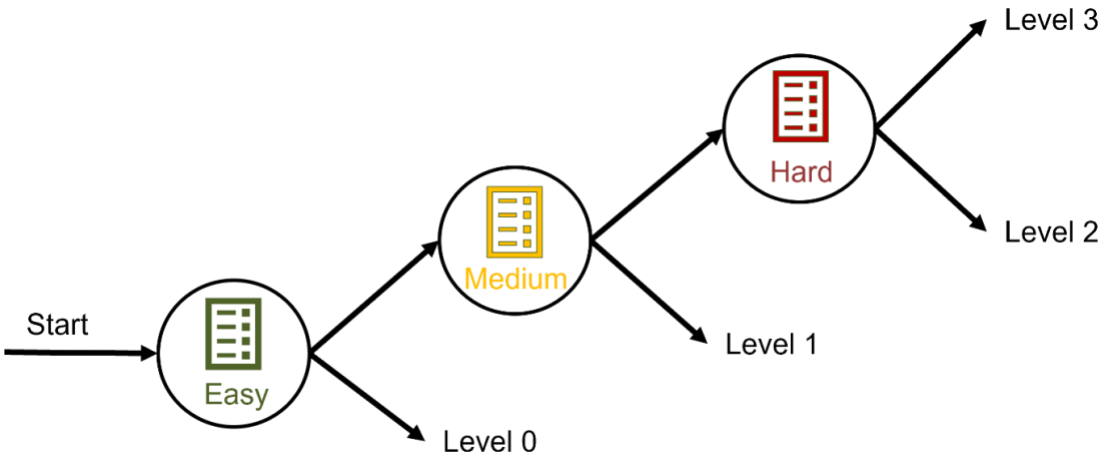

Figure 1: Difficulty levels in persona.

To assist students with their learnings, immediate feedback is adopted with persona in 4 aspects. First, for each question, there is generalised feedback to the students. Second, for each option for possible answers, specific feedback is provided to the students based on their answers. Third, the feedback connects the sections having the related contents to the attempted question. Lastly, the feedback relates the COs with the questions. Hence students get a clear idea about skills that are tested by a question.

The participants are added to the course website as students by giving them made-up names having generic usernames and password. The students can review the materials and attempt both formative assessments unlimited times. Finally, they complete 3 surveys: one for the effectiveness and usability of the traditional assessment, one for the effectiveness and usability testing of persona, and the last one is to compare the traditional assessment with persona.

The experimental study uses an explanatory sequential mixed methods design. Ethical approval for this study was granted by the UniSA Human Research Ethics Committee (Ethics Protocol 204940). Students from UniSA's

Bachelor of IT and Bachelor of Software Engineering were first invited via email to complete 3 surveys designed to capture their opinions on and experiences with online traditional assessment and persona. Focus groups with students were then conducted to further explain and expand upon questionnaire results (Creswell & Creswell, 2022).

Participant recruitment

Students were contacted by email in 2024-25 with the study participation invitation. The email contained an outline of the research project, a Participant Information Sheet and contact details of the research team to forward questions about the surveys or research project. Participation in the survey was voluntary. Consent from participants to be contacted for a follow-up focus group was obtained via the survey.

Table 1
SUS Questions

No.	Questions
1.	You think that you would like to use the online assessment frequently.
2.	You found the online assessment unnecessarily complex.
3.	You thought the online assessment was easy to use.
4.	You think that you would need the support of a technical person to be able to use this online assessment.
5.	You found the various tasks/functions in this online assessment were well integrated.
6.	You thought there was too much inconsistency in this online assessment.
7.	You would imagine that most people would learn to use this online assessment very quickly.
8.	You found this online assessment very boring/awkward to use.
9.	You felt very confident using the online assessment.
10.	You needed to learn a lot of things before you could get going with this online assessment.

Online survey questionnaires

The questionnaire included SUS questions (as shown in Table 1 above) to measure the usability of the online assessments. Usability was measured with the SUS scoring model (Lewis, 2018), shown in the following equation.

$$SUS = 2.5\left(20 + \sum_{i=1, i\ odd}^{10} SUS_i - \sum_{j=2, j\ even}^{10} SUS_j\right)$$

On the right-hand side of Equation 1, the SUS_i and SUS_j values where *i* and *j* denote the odd and even question numbers respectively (positive version) in Table 1. The responses ranged from 1 (strongly disagree) to 5 (strongly agree) for each of the questions. The adjective rating scale, which involved appending a 5-point Likert scale to the SUS. This question asked users to rate the user-friendliness of a system from the worst to the best imaginable.

In addition to SUS questions participants were also asked to respond to questions related to the effectiveness of the online assessments (e.g., responses captured in 5-point Likert scale).

4. Result and Discussion

Until 23 September 2025, 45 students have showed their interest to participate in the study and they have been added to the course website as students. Ten of these students never logged in to the course website. Out of the 35 students who visited the course website, 57.14% (n=20) fully completed the survey on traditional assessment, 54.29% (n=19) fully completed the survey on persona and 51.43% (n=18) fully completed the survey for comparing the traditional assessment with persona.

The age of participants spans between 18 years to 64 years, with around half of them are from 18-24 years range. The participation of female students is from 40.00% up to 44.44% for different surveys. Over 60% of the

participants have learned introductory level Python programming during the last 4 years (i.e., since 2022). The participants are from IT undergraduate programs namely IT(Generic), IT(Networking & Cyber Security), IT(Software Development) and IT(Games & Entertainment Design), with most from Bachelor of Software Engineering(Honours) (up to 35.00% in different surveys), followed by online Bachelor of Data Analytics (up to 33.33% in different surveys).

Usability Test

Overall, 20 students completed the SUS questions for traditional assessment, and 19 completed the SUS questions in relation to persona. The mean SUS score for the responses to traditional assessment was 76.9 (n=20), whereas for persona the mean was 74.2 (n=19), as shown below in Figure 2. According to (Bangor et al., 2009), this result indicates 'B' grade ("Good" usability) for both traditional assessment and persona meaning both traditional assessment and persona are acceptable and user-friendly. Students find both assessments usable.

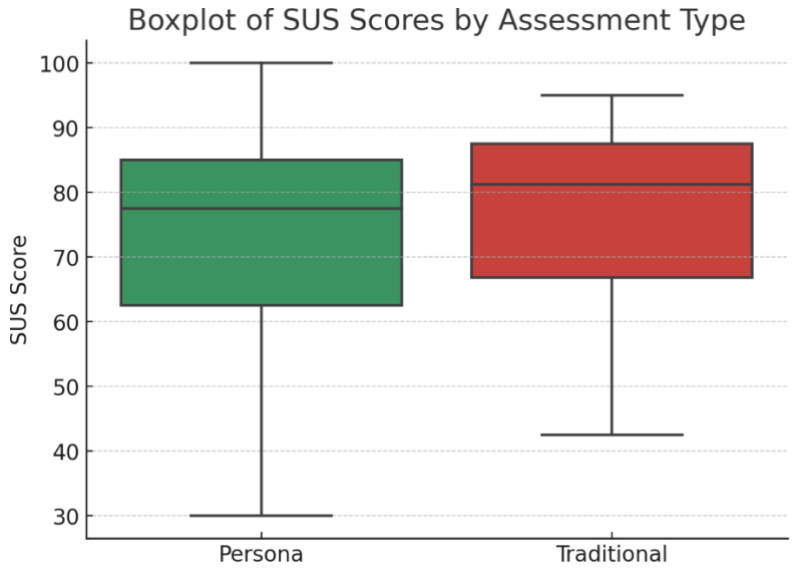

Figure 2: Boxplot showing the comparison of SUS scores of persona and traditional assessment.

Effectiveness

There were 3 questions for evaluating the effectiveness of these 2 online assessments, and these are related to supporting students' learning, alignment with COs, and helpful feedback. From Table 2 below, persona consistently rated higher in terms of learning support, alignment with COs, and helpful feedback compared to the traditional assessment.

Table 2
Descriptive Statistics - persona vs traditional assessment

Question	Persona Mean	Persona Median	Traditional Mean	Traditional Median
Q5 supporting students' learning	4.37	5.0	4.25	4.0
Q6 alignment with COs	4.37	5.0	4.20	4.0
Q7 helpful feedback	4.00	4.0	3.80	4.0

The diverging stacked bar charts clearly illustrate higher positive responses ("Agree" and "Strongly Agree") for persona, whereas more negative and neutral responses for traditional assessment as shown below in Figure 3.

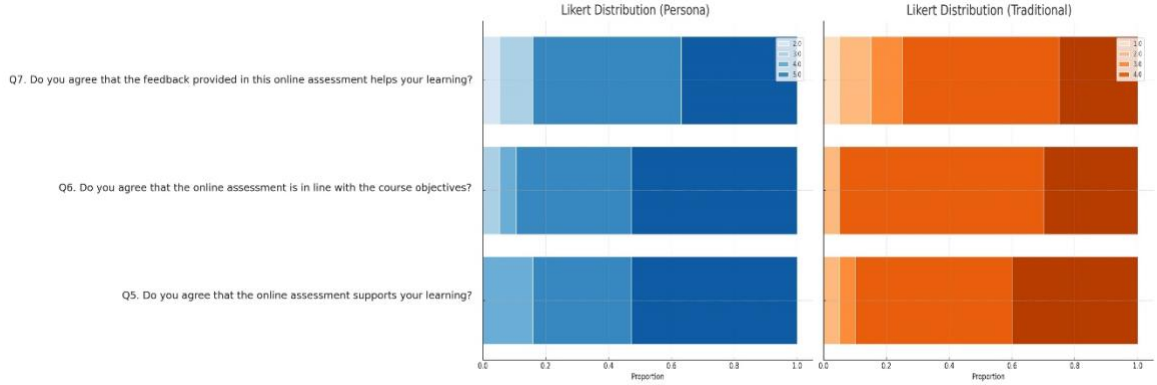

Figure 3: Likert Distribution

Table 3 does not show statistically significant differences between persona and traditional assessment (*p*-values range 0.29–0.62), even though descriptive statistics indicates somewhat higher ratings for persona.

Table 3
Mann–Whitney U Test (Statistical Differences) - persona and traditional assessment

Question	U-statistic	*p*-value
Q5 supporting students' learning	206.5	0.622
Q6 alignment with COs	224.0	0.292
Q7 helpful feedback	217.0	0.421

Comparison

Total 21 students attempted the comparison survey, of which 18 students completed the survey. Survey questionaries with 3 metrics are shown in Table 4 below, where column label '*For'* refers to strongly agree and agree combinedly and column label '*Against'* denotes strongly disagree and disagree collectively. The pattern of responses for the persona was more favourable for most of the criteria with few mixed experiences.

Out of 18 students, 72.2% would choose persona over traditional assessment and 77.8% would recommend persona to others. In both cases, traditional assessment performed poorly. Persona also performed far better in knowledge evaluation (66.7% showed confidence with persona's skills evaluation compared to 11.1% for traditional assessment) and in case of feedback, 72.2% students considered the feedback provided by persona had the potential to support their learning.

Although students expressed negative perception about passing the course by attempting a test using persona (about 38.9% did not go with persona), they showed positive perception regarding achieving higher grades by sitting a test using persona (over 55% showed confidence in persona). The persona received the worst negative attitude from the students with respect to the test length. Even though researchers like van der Linden & Glas (2000) stated that a test using CAT would take less time, here students experienced opposite. This might be because we set fixed time for the test using traditional assessment, while we did not set a timer for the test using persona.

5. Conclusion

In this study, usability and the effectiveness of online traditional assessment and persona were investigated from the perspectives of students, utilising results obtained from 3 questionnaires. Although the initial plan was to verify the questionnaire results through focus groups, we observed low interests for focus groups from students. As per quantitative results, persona significantly outperforms traditional assessment on supporting students' learning, alignment with COs, and helpfulness of feedback. We strongly consider maintaining and enhancing persona to optimise student experiences, students' learning and COs.

Table 4
Comparison between persona and traditional assessment

Criterion	For	Neutral	Against

When you take the test using personalised assessment you feel comfortable than when doing a test using traditional (i.e., non-personalised) assessment.	55.6	27.8	16.7
It is easier for you to pass the test using personalised assessment than a test using traditional (i.e., non-personalised) assessment.	44.4	16.7	38.9
It is easier for you to perform well (i.e. securing HD/D) in the test using personalised assessment than a test using traditional (i.e., non-personalised) assessment.	55.6	27.8	16.7
Taking the test using personalised assessment is less stressful than the test using traditional (i.e., non-personalised) assessment.	38.9	33.3	27.8
It takes less time for you to complete the test using personalised assessment than the test using traditional (i.e., non-personalised) assessment.	27.8	22.2	50.0
Your knowledge and skills in basic Python programming are evaluated more effectively and realistically by the test using personalised assessment, compared to the test using traditional (i.e., non-personalised) assessment.	66.7	22.2	11.1
The feedback provided by the test using personalised assessment has supported your learning better.	72.2	22.2	5.6
The questions having difficulty labels (e.g., easy, medium, hard) in the test using personalised assessment have helped you to attempt the test better than the unlabelled questions in the test using traditional (i.e., non-personalised) assessment.	55.6	16.7	27.8
If there is an open option to choose between personalised assessment and traditional (i.e., non-personalised) assessment, you will choose personalised assessment.	72.2	22.2	5.6
You will recommend personalised assessment for your peers.	77.8	16.7	5.6

In future, persona can be extended to map students' learning with respect to COs as shown below in Figure 4.

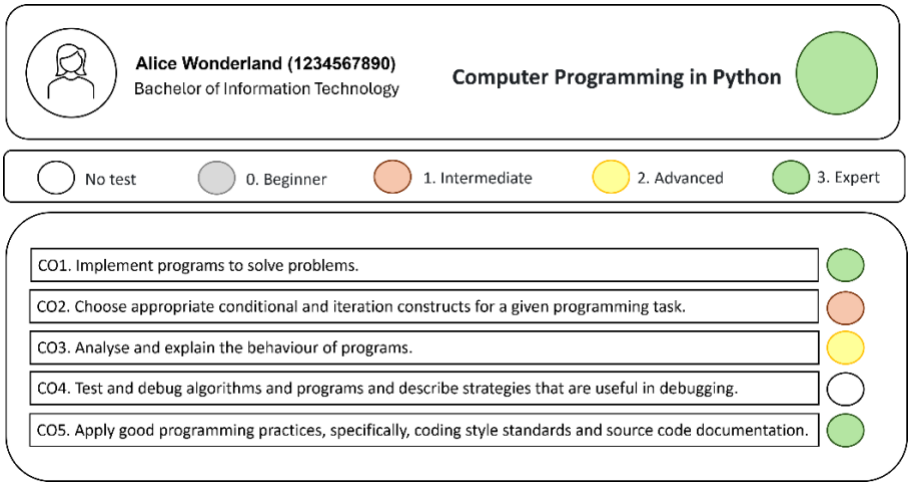

Figure 4: Conceptual dashboard showing overall course skill level as well as achievement levels for COs.

In addition to obtaining a score for a course or a topic, the students will receive scores for their ability for the COs. This will help students to know their status with respect to the COs and drive them to work on COs where they are behind. Furthermore, persona may incorporate personalised selection, where a student can choose the topic and the level that he or she wishes to be tested during a test using persona.

Acknowledgement

This project has been funded by UniSA Online Research Incentive Grant Scheme.

ASCILITE 2025

Future-Focused:

Educating in an Era of Continuous Change

References

ACARA, Australian Curriculum, Assessment and Reporting Authority (n.d.). Naplan. Retrieved July 7, 2025, from https://acara.edu.au/assessment/naplan.

ASCP BOC, American Society of Clinical Pathology Board of Certification (n.d.). Get Credentialed. Retrieved July 7, 2025, from https://www.ascp.org/content/board-of-certification/get-credentialed/#the_exam.

ASVAB, Armed Services Vocational Aptitude Battery. The CAT ASVAB. Retrieved July 7, 2025, from https://www.officialasvab.com/recruiters/cat-asvab/.

Bangor, A., Kortum, P., & Miller, J. (2009). Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of Usability Studies*, 4(3), 114–123. Online Journal Page.

Barker, T. (2010). An Automated Feedback System Based on Adaptive Testing: Extending the Model. *International Journal of Emerging Technologies in Learning* (*iJET*), 5(2), 11–14. https://doi.org/10.3991/ijet.v5i2.1235.

Black, P., & William, D. (2009). Developing the Theory of Formative Assessment. *Educational Assessment, Evaluation and Accountability*, 21, 5-31. https://doi.org/10.1007/s11092-008-9068-5.

Bennedsen, J., & Caspersen, M. E. (2007). Failure Rates in Introductory Programming. *ACM SIGCSE Bulletin*, 39(2), 32–36. https://doi.org/10.1145/1272848.127287.

Bennedsen, J., & Caspersen, M. E. (2019). Failure Rates in Introductory Programming: 12 Years Later. *ACM Inroads*, 10(2), 30–36. https://doi.org/10.1145/332488.

Chatzopoulou, D. I., & Economides, A. A. (2010). Adaptive assessment of student's knowledge in programming courses. *Journal of Computer Assisted Learning*, 26(4), 258-269. https://doi.org/10.1111/j.1365-2729.2010.00363.x.

Cisar, S. M., Cisar, P., & Robert Pinter, R. (2016). Evaluation of knowledge in Object Oriented Programming course with computer adaptive tests. *Computers & Education*, 92-93, 142-160. http://dx.doi.org/10.1016/j.compedu.2015.10.016.

Creswell, J. W., & Creswell, J. D. (2022). Research design: Qualitative, quantitative, and mixed methods approaches (5th edition). SAGE Publications.

Csapo, B., & Molnar, G. (2019). Online diagnostic assessment in support of personalized teaching and learning: The eDia system. *Frontiers in Psychology*, 10(1522), 1-14. https://doi.org/10.3389/fpsyg.2019.01522.

Eggen, T. J. H. M. (2007). Choices in CAT models in the context of educational testing. In D. J. Weiss (Ed.), Proceedings *2007 GMAC Conference on Computerized Adaptive Testing*. Retrieved July 7, 2025, from https://scispace.com/pdf/choices-in-cat-models-in-the-context-of-educational-testing-3pgp3jnnve.pdf.

Evans, C. (2013). Making sense of assessment feedback in higher education. *Review of Educational Research*, *83*(1), 70–120. https://doi:10.3102/0034654312474350.

Franco, A., & Potenko, V. (2022), *Adaptive Quiz* (1.2.3 version). [Moodle LMS Plugin]. https://moodle.org/plugins/mod_adaptivequiz.

Gierl, M., Bulut, O., & Zhang, X. (2018). Using Computerized Formative Testing to Support Personalized Learning in Higher Education: An Application of Two Assessment Technologies. In R. Zheng (Ed.), *Digital Technologies and Instructional Design for Personalized Learning* (pp. 99-119). IGI Global Scientific Publishing. https://doi.org/10.4018/978-1-5225-3940-7.ch005.

GMAC, Graduate Management Admission Council (2025, May 23). GMAT: What Is a Computer-Adaptive Exam? Retrieved July 7, 2025, from https://support.mba.com/hc/en-us/articles/7261613974043-GMAT-What-Is-a-Computer-Adaptive-Exam.

Hattie, J. (2009). Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement. Routledge. https://doi.org/10.4324/9780203887332.

Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, 77(1), 81–112. https://doi.org/10.3102/003465430298487.

Ihichr, A., Oustous, O., El Idrissi, Y. E. B., & Lahcen, A. A. (2024). A Systematic Review on Assessment in Adaptive Learning: Theories, Algorithms and Techniques. *International Journal of Advanced Computer Science & Applications*, *15*(7), 855-868. 10.14569/IJACSA.2024.0150785.

Jacobsen, J., Ackermann, R., Egüez, J., Ganguli, D., Rickard, P., & Taylor, L. (2011). Design of a computer-adaptive test to measure English literacy and numeracy in the Singapore workforce: Considerations, benefits, and implications. *Journal of Applied Testing Technology*, 12(special4). Association of Test Publishers. https://eric.ed.gov/?id=EJ943067.

Krathwohl, D. R. (2002). A Revision of Bloom's Taxonomy: An Overview. *Theory Into Practice*, *41*(4), 212–218. http://www.jstor.org/stable/1477405.

Lewis, J. R. (2018). The System Usability Scale: Past, Present, and Future. *International Journal of Human-Computer Interaction, 34*(7), 577–590. https://doi.org/10.1080/10447318.2018.1455307.

Lilley, M. (2007). The development and application of computer adaptive testing in a higher education environment. *PhD thesis*. Retrieved July 7, 2025, from https://uhra.herts.ac.uk/id/eprint/16174/.

Lilley, M, Barker, T., & Britton, C. (2004). The development and evaluation of a software prototype for computer-adaptive testing, *Computers & Education*, 43(1–2), 109-123. https://doi.org/10.1016/j.compedu.2003.12.008.

Medeiros, R. P., Ramalho, G. L., & Falcao, T. P. (2019). A Systematic Literature Review on Teaching and Learning Introductory Programming in Higher Education. *IEEE Transactions on Education*, 62(2), 77-90. 1109/TE.2018.2864133.

Mohorovicic, S., & Strcic, V. (2011). An Overview of Computer Programming Teaching Methods. Proceedings *22nd Central European Conference on Information and Intelligent Systems*, Faculty of Organization and Informatics Varazdin, Croatia (pp. 47–52). Document URL on ProQuest.

Olea, J., Abad, F. J., Ponsoda, V., Aguado, D., & Díaz, J. (2012). Development, psychometric properties and new validity evidences of the web-based computerized adaptive test of English eCat. *R.E.M.A. Revista electrónica De metodología Aplicada*, *16*(1), 50–65. https://reunido.uniovi.es/index.php/Rema/article/view/9796.

Olea, J., Abad, F. J., Ponsoda, V., Barrada, J. R., & Aguado, D. (2011). eCAT-Listening: design and psychometric properties of a computerized adaptive test on English Listening. *Psicothema*, *23*(4), 802–807. https://pubmed.ncbi.nlm.nih.gov/22047876/.

Pardo, A., Bartimote, K., Buckingham Shum, S., Dawson, S., Gao, J., Gašević, D., Leichtweis, S., Liu, D., Martínez-Maldonado, R., Mirriahi, N., Moskal, A. C. M., Schulte, J., Siemens, G., & Vigentini, L. (2018). OnTask: Delivering Data-Informed, Personalized Learning Support Actions. *Journal of Learning Analytics*, *5*(3), 235–249. https://doi.org/10.18608/jla.2018.53.15.

Resnick, M., Maloney, J., Monroy-Hernández, A., Rusk, N., Eastmond, E., Brennan, K., Millner, A., Rosenbaum, E., Silver, J., Silverman, B., & Kafai, Y. (2009). Scratch: Programming for all. *Communications of the ACM*, 52(11), 60–67. https://doi.org/10.1145/1592761.1592779.

Schonenberg, N., Van Achter, I., Van Walle, A., Van Deun, K., Decoo, W., & Colpaert, J. (1993). ATLAS: An Adaptive Model for Measuring Language Proficiency. *CALICO Journal*, *11*(2), 45–50. http://www.jstor.org/stable/24147573.

Stone, C., & Crawford, N. (2020). Three Essentials in Moving Online. *National Centre for Student Equity in Higher Education* (*NCSEHE*), Curtin University. https://www.ncsehe.edu.au/wp-content/uploads/2020/04/3-Essentials-online.pdf.

TTBJ, Tsukuba Test – Battery of Japanese (n.d.). University of Tsukuba. Retrieved July 7, 2025, from https://ttbj.cegloc.tsukuba.ac.jp/.

UCF (2011). Design and Delivery Principles. *Blended Learning Toolkit*, University of Central Florida. https://blended.online.ucf.edu/2011/06/07/design-delivery-principles/.

UNSW eLearning (2014). *U21 Conference - Keynote - Simon Bates "Personalised Learning"* [Video]. YouTube. https://www.youtube.com/watch?v=wFetEo5qCqc.

Van der Linden, W. J, & Glas, G. A. W. (eds) (2000). Computerized Adaptive Testing: Theory and Practice. *Springer Dordrecht*. https://doi.org/10.1007/0-306-47531-6.

Watson, C., & Li, F. W. B. (2014). Failure rates in introductory programming revisited. Proceedings *Conference on Innovation and Technology in Computer Science Education* 2014 (ITiCSE '14), Uppsala, Sweden (pp. 39–44). https://dl.acm.org/doi/10.1145/2591708.2591749.

Karim, S. M. M., Mirzaei, S., Hossain, S. N., Abadia, R. & Islam, A. (2025). Scaffolding students' learning of introductory programming in online higher education by personalised formative assessments. In S. Barker, S. Kelly, R. McInnes & S. Dinmore (Eds.), *Future-focused: Educating in an era of continuous change*. Proceedings ASCILITE 2025. Adelaide (pp. 231-240). https://doi.org/10.65106/apubs.2025.2643